PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68 — A2

(11) International Publication Number: **WO 00/29621**

(43) International Publication Date: 25 May 2000 (25.05.00)

(21) International Application Number: PCT/US99/26931

(22) International Filing Date: 12 November 1999 (12.11.99)

(30) Priority Data:
09/193,320 16 November 1998 (16.11.98) US

(71) Applicant: GENELABS TECHNOLOGIES, INC. [US/US]; 505 Penobscot Drive, Redwood City, CA 94063–4738 (US).

(72) Inventors: DOLGANOV, Gregory; 14 Chicory Lane, San Carlos, CA 94070 (US). NOVIKOV, Alexander; 5866 Carmel Way, Union City, CA 94587 (US).

(74) Agents: SUH, Young, J.; Legal Department, Genelabs Technologies, Inc., 505 Penobscot Drive, Redwood City, CA 94063–4738 (US) et al.

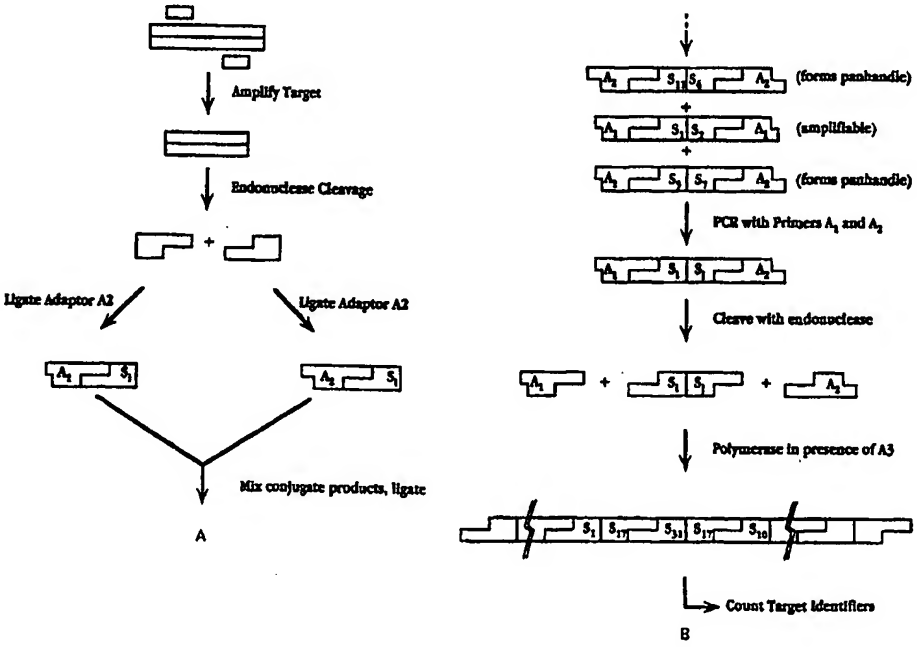
(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD FOR MEASURING TARGET POLYNUCLEOTIDES AND NOVEL ASTHMA BIOMOLECULES

(57) Abstract

The present invention provides a method for simultaneously determining the levels of selected target polynucleotide sequences in a sample. In the method, selected target sequences are amplified using sequence–selective primer pairs to form double–stranded copies. The copies are cleaved with one or more endonucleases, and first and second aliquots of the cleaved fragment mixture are ligated to first and second adaptors, respectively, to form mixtures of first and second adaptor–fragment conjugates. The aliquots are combined and ligated to form conjugate dimers containing a conjugate from each of the first and second conjugate mixtures. After optional amplification, the conjugates are treated to release the adaptor segments, yielding target–identifier dimers. The dimers are polymerized to form target–identifier dimer multimers, and the relative abundances of the target–identifiers in one or more dimer multimers to provide an estimate of the levels of the selected target sequences in the sample. The present invention is particularly useful for determining expression levels of mRNA gene sequences in a sample, and for evaluating changes in mRNA expression levels among different samples or in response to changes in sample conditions. Also disclosed are certain polynucleotides and polypeptides that are useful for a variety of applications, particularly relating to asthma.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHOD FOR MEASURING TARGET POLYNUCLEOTIDES AND NOVEL ASTHMA BIOMOLECULES

Field of the Invention

The present invention relates to methods for measuring target polynucleotides in samples. In a preferred embodiment, the invention also relates to characterizing expression levels of genes, preferably for quantifying expression levels of selected mRNA species from selected tissue and cellular samples. The invention also includes certain novel polynucleotides and polypeptides which are up-regulated in asthma patients, and methods and compositions relating thereto.

Introduction

Many diseases have been linked to specific genetic defects such as polymorphisms, deletions, insertions, and post-replication mutations. Such defects have been associated with specific physiological changes, such as the functions and morphologies of cells and tissues, and in some cases, have been traced to underlying changes in the structures or catalytic activities of proteins encoded by the altered sequences. Expression levels of mRNAs encoded by genomic DNA are also very important in physiological processes. It has been estimated that the human genome encodes approximately 100,000 genes (Fields, C., *et al.* (1994) Nat. Genet. 7:345-346). However, only a select subset of genes are expressed in a cell at any given time, depending on the cell type, developmental stage, age, cellular stress, and other factors. For example, the expression profile of mRNAs in a leukocyte will differ from the expression profile observed in an endothelial cell or red blood cell. Likewise, the profile of a healthy cell will differ from that of the same cell type infected by a virus or bacterium. The non-expression of a required gene, or the improper expression of a gene, can lead to serious or fatal consequences. Moreover, an increase in the level of expression of a gene by 50% over that of the normal level can lead to a disease condition (Audic *et al.* (1997) Genome Res. 7:986-995).

Functional genomics refers to a new phase of genome analysis concerned with the development and application of global experimental approaches to assess gene function. Current sequencing efforts on the human genome are yielding vast amounts of sequence information, but processing and interpreting the data is not straightforward. Functional

roles of genes are predicted on the basis of similarities with sequences of known function, which are usually from other organisms. However, such predictions are usually qualitative and are incapable of discerning a gene's true activities and functions.

Often, information is available about a particular gene that suggests possible or likely involvement in a disease condition or other biological process, but substantial further research is required before the nature of the involvement is clear. For example, genetic linkage studies can reveal that a gene or group of genes is associated with a particular disease and warrants further investigation. As part of characterizing the function of such genes, it is often desirable to determine whether and to what extent the genes of interest are expressed under normal and abnormal conditions.

There is currently a need for a convenient method for specifically and accurately measuring expression levels of genes of interest, especially to correlate expression levels with biological activity. More generally, it would be desirable to develop a method that allows rapid, simultaneous quantification of different target sequences in a polynucleotide sample. Such methods promise to be highly useful in detecting and monitoring biological conditions, and for discovering new polynucleotides and polynucleotides of diagnostic and therapeutic value. In addition, such a method would be useful for characterizing genetic sequences and gene expression associated with many diseases, such as asthma, cardiovascular diseases, hepatitis, etc.

As an example, it is now widely accepted that asthma in its varied forms is an inflammatory disorder of the airways in which mediator release from activated mast cells and eosinophils plays a major role. Asthma is defined by three characteristic features: intermittent reversible airway obstruction, airway hyperresponsiveness, and airway inflammation. Although all of these components are important, it is generally accepted that the primary underlying abnormality in this disease is a unique form of chronic inflammation. The histopathological hallmark of bronchial asthma, even in the mild and intermittent syndromes, is the constant presence of epithelial lesions in the bronchial mucosa, thickening of basement membranes, and inflammatory infiltration consisting of activated eosinophils, Th2 lymphocytes and degranulated mast cells. T lymphocytes take a primary role in orchestrating these processes through their capacity to generate a range of cytokines of the interleukin 4 gene cluster encoded on the long arm of chromosome 5.

The attraction of leukocytes to tissues is essential for inflammation and immune response in asthma/atopy. This process is controlled by the coordinated production of

chemokines, which are a large family of structurally related small basic proteins. These proteins regulate migration and activation of leukocytes through G protein-coupled cell-surface receptors. To date, several dozen chemokines have been identified which have been subdivided into at least four different families (reviewed by Krug, N. *et al.* (1998) Clin. Exper. Allergy 28:10-18; Luster A.D. *et al.* (1998) N. Engl. J. Med. 338(7): 436-445), based in part on the total number of cysteines in the molecule and the spacing between the two most N-terminal cysteines. The most extensively characterized chemokines belong to the alpha- and beta-chemokine families, designated CXC and CC chemokines, respectively. Chemokines and their receptors are differentially expressed on the different subsets of T helper cells, suggesting that regulated networks of gene expression may control tissue-specific migration of T helper cells.

It is an object of the invention to provide a method that is applicable to any known nucleic acid sequence of interest. A further object is to provide a method that allows rapid, simultaneous measurement of a plurality of target sequences of interest with high sensitivity and accuracy. Yet another object is to provide a method that allows sensitive measurement of low copy number mRNA species, including species with average expression levels below one copy per cell.

It is a further object of the invention to provide polynucleotides which encode novel polypeptides. The expression of these novel polypeptides, measured using methods provided herein, have been discovered to be significantly up-regulated in tissues and body fluids of asthma patients.

It is further an object to provide polynucleotides and novel polypeptides whose levels are elevated in asthma, and methods and compositions relating thereto.

Summary of the Invention

The present invention provides, in one aspect, a method for simultaneously determining the levels of a plurality of selected target polynucleotide sequences in a sample. In the method, the sample is reacted with a plurality of sequence-selective primer pairs, wherein each primer pair is designed to flank a defined region in a different target sequence of interest, and each defined region contains a selected restriction endonuclease site, to form double-stranded copies of the defined regions in direct proportion to the target levels originally present in the sample. The double-

stranded copies are then cleaved with one or more endonucleases, to form a cleavage mixture of DNA fragments.

First and second aliquots of the cleavage mixture are reacted separately with first and second adaptors, respectively, to form mixtures of first and second adaptor-fragment conjugates. Preferably, the first and second adaptors each contain (i) a first terminal end that is cohesive with the cleaved termini in the DNA fragments, (ii) a second terminal end that is non-cohesive with respect to the DNA fragments, so that only one end of each adaptor binds to a fragment in the DNA fragment mixture, and (iii) a unique-sequence segment that allows the first and second adaptors to be distinguished from each other. The resultant conjugates comprise an adaptor segment and a target-identifier segment. Optionally, one or more of the adaptors contains a binding site for a Type IIs restriction endonuclease so that the cleavage site is located within the target segment, and the conjugate products in the first and second conjugate mixtures are cleaved with the corresponding Type IIs restriction endonuclease to shorten the target-identifier segments in the conjugates prior to dimerization.

The first and second conjugate mixtures are combined under conditions effective to form conjugate dimers containing a conjugate from each of the first and second conjugate mixtures, such that the conjugates in each dimer are joined via their target-identifier segments and are flanked by the adaptor segments. The adaptors can then be cleaved from the conjugate dimers, typically by treatment with at least one endonuclease, which may be the same or different from the first mentioned endonuclease, to produce a plurality of target-identifier dimers. Optionally, these dimers are amplified by one or more cycles of polymerase chain reaction.

The dimers are then polymerized to form dimer multimers. From the relative abundances of target-identifiers in one or more dimer multimers, the expression levels of the selected target sequences can be determined. In one embodiment, the multimers are inserted into a cloning vector to form a multimer library, from which any number of different multimers can be sequenced to determine target abundances with the desired level of accuracy and precision.

Preferably, the target-identifiers in the multimers have lengths of from 10 to 50 base pairs, and more preferably from 10 to 20 base pairs. Also, the multimers preferably contain from 2 to 100 target-identifier dimers, and preferably 20 to 60 target-

identifier dimers. Typically, sequenced multimers contain from 20 to 2000 base pairs, preferably at least 400 base pair, typically 400 to 800 base pairs.

In a particularly preferred embodiment, the double-stranded copies that are formed have substantially identical lengths, by use of appropriately designed primer pairs.

The sample oligonucleotides may be single or double-stranded. For example, the target polynucleotides can be single or double-stranded DNA molecules. In another embodiment, the selected target polynucleotides are RNA molecules, preferably mRNA molecules.

In a particularly preferred embodiment, the defined target regions are gene sequences encoded by mRNA, and the double-stranded copies that are formed are cDNA fragments. In one exemplary embodiment, first strand cDNA synthesis can be accomplished by contacting a sample mixture of mRNAs with a plurality of gene-specific primers in the presence of a reverse transcriptase, under conditions effective to form, by primer extension, a plurality of mRNA/DNA hybrids comprising first cDNA strands. The cDNA strands can then be contacted with a plurality of second primers from the above-mentioned primer pairs, in the presence of a DNA polymerase, under conditions effective to form a plurality of double-stranded cDNAs by primer extension. Preferably, the double stranded cDNAs are amplified simultaneously by at least one cycle of polymerase chain reaction in the presence of a plurality of the above primer pairs. In another exemplary embodiment, the first strand synthesis is performed using a first primer to form a complementary cDNA strand, and further amplification is performed using a primer pair comprising second and third primers for amplification, wherein the first primer is binds to a sequence in the target polynucleotide that is outside the target binding sequences of the second and third primers. In a third general embodiment, targets (DNA or RNA) are amplified by nested PCR, using a first primer set comprising first and second primers for one or more initial amplification cycles, followed by amplification using a second primer set comprising third and fourth primers which amplify a sub-sequence within the sequence amplified by the first primer set.

In another aspect, the invention includes a method for simultaneously determining expression levels of a plurality of selected target gene sequences in a sample. In one embodiment, a cDNA fragment mixture is formed by treating the sample with a plurality of gene-specific primer pairs, wherein each pair comprises first

and second primers which are complementary to opposite ends of a different selected region of a selected target gene, and each selected region encodes a selected restriction endonuclease site. The double-stranded copies may be cleaved with one or more endonucleases, to form a cleavage mixture of cDNA fragments, and at least two aliquots of the mixture are reacted separately with different adaptors, to form adaptor-fragment conjugates. The mixtures are combined, and ligation, cleaving, polymerizing, and quantifying steps may be performed as described herein.

In yet another aspect, the invention includes a polynucleotide mixture comprising a plurality of dimer multimers prepared as herein. Preferably, the target-identifiers in the multimers have lengths of from 10 to 50 base pairs, and more preferably from 10 to 20 base pairs. Also, the multimers preferably contain from 2 to 100, and preferably 20 to 60 target-identifier dimers. Typically, the multimers contain from 20 to 2000, and preferably from 400 to 800 base pairs. In a particularly preferred embodiment, the target-identifier segments in the multimers have substantially identical lengths. Preferably, the target-identifier segments are derived from mRNA gene sequences, in which case, the target-identifier can also be referred to as a gene-identifier.

In yet another embodiment, the invention includes a method of determining the effects of environmental conditions on the levels of one or more selected target polynucleotide species in a sample. Preferably, the target species are mRNAs. In the method, levels of a plurality of different target polynucleotides are measured in a sample by any of the methods above under a first set of environmental conditions, and then the levels are measured again under a second set of environmental conditions which are different from the first conditions. Any observed change in transcription level, or the lack of such change, can be used to characterize the consequences of the test conditions. In a preferred embodiment, the method is useful as a tool for screening and characterizing drugs and drug candidates for ameliorative and/or deleterious affects on gene expression.

The present invention further provides polynucleotides and polypeptides that are useful for a variety of applications, particularly relating to asthma.

In one aspect, the invention includes an isolated and purified polypeptide, designated herein as P2, having an amino acid sequence at least 80% identical to SEQ ID NO:2 or a splice isoform SEQ ID NO:4. In other embodiments, the polypeptide has a sequence at least 85%, at least 90%, at least 95%, or at least substantially identical to

SEQ ID NO:2 or SEQ ID NO:4. The invention also includes fragments of P2, which are antigenic or which are capable of interacting with other proteins, peptides, or chemicals, such interaction which alters the functional properties or cellular/subcellular localization of P2 or a P2 receptor.

In another aspect the invention includes an isolated polynucleotide having a sequence which encodes P2 as described above, or a sequence complementary to the P2 coding sequence, and a composition comprising the polynucleotide. The polynucleotide may be mRNA, cRNA, DNA, cDNA, genomic DNA, or an antisense analog thereof. In various embodiments the nucleic acid may encode a P2 having an amino acid sequence at least 80% identical to SEQ ID NO:2. In a general embodiment, the polynucleotide has at least 70%, preferably at least 80%, 90%, or 95% sequence identity with the P2 sequence identified as SEQ ID NO:1 or the isoform SEQ ID NO:3, or with the open reading frames (ORFs) corresponding to nucleotides 192-422 of SEQ ID NO:1, or nucleotides 192-311 of SEQ ID NO:3. In other embodiments, the polynucleotide has at least 70%, preferably at least 80%, 90%, or 95% sequence identity with one or more of the exons or introns as identified in Section II.B1 below, either singly or in combination, and, if in combination, joined together in consecutive or non-consecutive order. In other embodiments, the polynucleotide has a sequence (a) essentially identical to any of the sequences described above, or (b) to its complement, or (c) hybridizes under at least high-stringency conditions to (a) or (b).

The invention also contemplates polynucleotides at least 12 nucleotides in length, preferably at least 15 nucleotides in length, more preferably at least 20, 25, 30, or 50 nucleotides in length, which hybridize under at least high-stringency conditions to a P2 nucleic acid sequence described above. The polynucleotide may be mRNA, cRNA, DNA, cDNA, genomic DNA, or an antisense analog thereof.

Also disclosed is a recombinant expression vector containing a polynucleotide encoding P2 as described above, and, operably linked to the polynucleotide, regulatory elements effective for expression of the polypeptide in a selected host. Preferred coding sequences are given above. In a related aspect, the invention includes a host cell, preferably a eukaryotic host cell, containing the vector.

The invention further includes a method for producing P2 by recombinant techniques, by culturing recombinant host cells containing a nucleic acid encoding P2

under conditions promoting expression of the protein, and subsequent recovery of the protein from the host cell.

In still another aspect, the invention includes an antibody specific for P2. The antibody has diagnostic and therapeutic applications, particularly in treating inflammation and asthma. Treatment methods which employ antisense or coding sequence polynucleotides for inhibiting or enhancing levels of P2 are also contemplated, as are treatment methods which employ antibodies specific for P2.

Diagnostic methods for detecting levels of P2 in specific tissue samples, and for detecting levels of expression of P2 in tissues, also form part of the invention. In one embodiment, a method of detecting a polynucleotide which encodes P2 in a biological sample, involves the steps of: (a) hybridizing a polynucleotide, which is capable of hybridizing to a polynucleotide which encodes P2, to nucleic acid material of a biological sample, thereby forming a hybridization complex, and (b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of the polynucleotide encoding P2 in the biological sample. Methods for detecting mutations in the gene encoding P2 are also contemplated.

Screening methods which employ P2 for identifying a candidate compound which modulates the activity of P2 also form part of the invention. An exemplary method includes (a) contacting a test compound with P2, under conditions in which an activity of P2 can be measured, (b) measuring the effect of the test compound on the activity of P2, and (c) identifying the test compound as a candidate compound if its effect on the activity of P2 is above a selected threshold level. The activity measured may be, for example, the expression level of a P2 polypeptide, the expression level of a P2-encoding polynucleotide, or the binding of P2 to a P2 receptor. In one embodiment, the test compound is a component of a combinatorial library. In another embodiment, the test compound is an antibody specific for P2.

The invention also includes, in a related aspect, a compound identified by the screening methods described above, including a purified agonist and a purified antagonist. The invention further includes a purified antibody which specifically binds to a polypeptide described above.

The invention also includes methods to alter the expression level of P2 by gene therapy techniques to achieve therapeutic benefits in patients.

These and other objects and features of the invention will become more fully apparent when read in light of the appended drawings and sequence listing and the following detailed description.

Brief Description of the Drawings

Figs. 1A-1B illustrate a general embodiment of the present invention;

Fig. 2 illustrates a panhandle structure formed by denatured single strands from an adaptor-fragment conjugate dimer which terminates at each end with the same type of adaptor;

Fig. 3 illustrates exemplary double-stranded adaptors according to Example 4.

Fig. 4 illustrates an embodiment in which adaptor-fragment conjugates formed in the method are trimmed with a type II restriction endonuclease to shorten the gene fragment to be measured;

Fig. 5 shows a plot of mRNA expression levels measured in bronchoalveolar lavage samples from healthy and diseased subjects for several selected genes;

Fig. 6 shows a plot of mRNA expression levels measured in blood samples from healthy and diseased subjects for several selected genes; and

Fig. 7 shows an amino acid sequence alignment between P2 from human (SEQ ID NO:2) and P2 from mouse (SEQ ID NO:7) prepared using the CLUSTAL-W (ver. 1.4) alignment program of MACVECTOR (Oxford Molecular Ltd. Oxford, UK) in pairwise mode, using the default pairwise alignment parameters (Open gap penalty = 10.0, Extend gap penalty = 0.1, and Blosum30 Similarity Matrix).

Brief Description of the Sequences

SEQ ID NO:1 is a nucleic acid sequence encoding a human P2 polypeptide;

SEQ ID NO:2 is the predicted translation product of SEQ ID NO:1;

SEQ ID NO:3 is a nucleic acid sequence encoding an exemplary human P2 splice isoform designated P2-B;

SEQ ID NO:4 is the predicted translation product of SEQ ID NO:3;

SEQ ID NO:5 is a human genomic sequence of P2;

SEQ ID NO:6 is a nucleic acid sequence encoding a mouse P2 polypeptide;

SEQ ID NO:7 is the predicted translation product of SEQ ID NO:6;

SEQ ID NO:8 is the nucleic acid sequence of the forward strand of exemplary Adaptor 1 described in Example 4;

SEQ ID NO:9 is the nucleic acid sequence of the reverse strand of Adaptor 1;

SEQ ID NO:10 is the nucleic acid sequence of the forward strand of Adaptor 2;

SEQ ID NO:11 is the nucleic acid sequence of the reverse strand of Adaptor 2;

SEQ ID NO:12 is the nucleic acid sequence of the forward primer described in Example 4;

SEQ ID NO:13 is the nucleic acid sequence of the reverse primer described in Example 4; and

SEQ ID NO:14 is the nucleic acid sequence of the PDS1-CUA primer described in Example 4.

SEQ ID NO:15 and 16 are constituent first and second strands of terminator adaptor A3 utilized in Example 4.

SEQ ID NO:17 through 50 represent exemplary primer sets for amplifying 17 selected target sequences (see Appendix I and Figs. 5 and 6).

Detailed Description of the Invention

As discussed in section I below, the present invention provides a rapid method for simultaneously determining absolute and/or relative levels of different-sequence polynucleotides in a sample. By using sequence-selective primers to generate short segments from the selected targets, the levels of specific target species can be measured to the exclusion of large numbers of other polynucleotides in the sample. The methods of the invention are conveniently applicable to a wide variety of sample types, and can be used to quantify low copy number targets with high sensitivity and specificity. As illustrated further herein, the invention can also be used to identify new markers of disease that can be adapted to diagnostic testing, health monitoring, and drug screening. In addition, the invention can be used to investigate mechanisms of action of various chemical reagents and biochemical conditions and events that may interact with expression of mRNA.

Section II describes another aspect of the invention relating to novel polypeptides, designated herein as P2 polypeptides, that are up-regulated in asthma patients, and polynucleotides which encode the P2 polypeptides. In one embodiment,

the P2-encoding polynucleotides can be monitored using the methods described in Section I.

I. Polynucleotide Measurements

The present invention provides, in one aspect, a method for simultaneously determining the levels of a plurality of selected target polynucleotide sequences in a sample.

The sample can be any sample that contains, or possibly contains, target polynucleotides of interest. Exemplary but non-limiting examples include: eukaryotes and prokaryotes; bacteria such as E. coli, B. subtilis, streptococcus, etc.; fungi such as yeast; animals, particularly mammals such as humans, rats, mice, sheep, cows, horses, dogs and cats; and plants, particularly food crops such as corn, wheat, and rice, for example. The samples may be derived from such sources in any suitable manner, and may take a variety of forms, such as tissues, organs, cells, extracts, homogenates, lysates, bronchoalveolar lavages, needle biopsies, and the like. The sample can also be derived from a selected cell population or cell type, such as endothelial cells, fibroblasts, bone cells such as osteoblasts and osteoclasts, blood cells such as red blood cells and white blood cells, and from organelles such as mitochondria, for example. The sample can be selected or modified to be in a particular developmental stage (e.g., a progenitor cell, a differentiated cell, or a cell undergoing division or apoptosis), or to be in a selected state (e.g., healthy, diseased, infected, etc.).

Figure 1A-1B shows an overview of an exemplary embodiment of the invention, starting from double-stranded DNA. In the first step, a sample is reacted with a plurality of sequence-selective primer pairs to form double-stranded copies of selected target polynucleotides containing defined target sequences. Here, "sequence-selective" means that each primer successfully hybridizes to the intended complementary target sequence. Perfect complementarity between primers and targets is preferred but not essential. Also, primers may contain degenerate bases at one or more positions to ensure priming to different possible sequence variants.

The target polynucleotides can be RNA or DNA, and may be single or double stranded. RNA targets can include mRNA, tRNA, snRNA, rRNA, or any other class of RNAs.

Target polynucleotide sequences are selected by the user, based on known sequence information. For example, in a sample that contains a mixture of different viruses, primer pairs can be designed to amplify viral sequences that are unique to each strain. Similarly, for measuring mRNA levels, primer pairs can be designed to amplify mRNA species for known or putative gene products whose expression levels are of interest. For example, target mRNAs can include mRNAs that encode various cytokines, such as the interleukins, growth factors, and the like, as well as known or putative gene transcripts that are localized in chromosomal gene clusters.

The double-stranded DNA can be prepared from any suitable source, such as total DNA, genomic DNA, plasmid DNA, mRNA, or the like, by any method known in the art (e.g., Sambrook et al., Molecular Cloning, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1990). mRNAs may be prepared by any known method, e.g., by poly(dT) affinity chromatography for eukaryotic mRNAs, etc. Preferably, mRNA is isolated from the sample as a total RNA fraction, to increase the recovery of low abundance mRNAs. Also, RNA samples can be treated with DNAse to break down DNA strands that might otherwise act as templates in the presence of DNA polymerases.

In a preferred embodiment of the invention, for measuring expression levels of selected target genes, mRNAs can be utilized as templates to form double-stranded cDNAs. First-strand cDNA synthesis is performed on mRNA templates using selected primers and a reverse transcriptase according to standard synthesis conditions. Reverse transcriptases and conditions for their use are known in the art and are readily available through commercial sources. The reverse transcriptase employed in the present invention preferably has high fidelity base incorporation to minimize sequence errors. The mRNA sample is reacted with reverse transcriptase, dNTPs, and appropriate synthesis primers such that first DNA synthesis strands encompassing the target sequences of interest are formed preferably on most or essentially all of the mRNAs in the sample. However, it is not essential that first strand synthesis be accomplished on all available mRNA templates, provided that the relative ratios of the target mRNAs are preserved to the extent necessary to provide satisfactory data.

In a preferred mode of the invention, the primers are selected to be complementary to target gene sequences of interest, such that first complementary DNA strands are formed only on mRNAs that contain the genes of interest. One advantage of

gene-specific primers is that background levels of other, non-target gene sequences are significantly reduced. Also, the proximity of each gene-specific primer to its corresponding selected target gene sequence helps ensure that the synthesized cDNA strands extend through and beyond the sequences of interest, even if extension along a particular mRNA template ceases before reaching the 5'-end.

In another approach, the selected primers for first cDNA strand synthesis are poly-T oligonucleotide primers designed to initiate extension at the 3'-ends of all mRNAs. The poly-T primers preferably comprise three distinct primers, each of which terminates with a 3'-A, C, or G nucleotide to anchor the primer to all possible poly-A tail/non-poly-A tail junctions in the mRNAs, to improve priming efficiency. Poly-T primers may be used to form first-strand cDNAs for all mRNAs in the sample, followed later by gene-specific amplification using gene-specific primer pairs to enrich for the target gene sequences of interest.

Double-stranded cDNAs are formed by annealing a second, gene-specific primer to the first cDNA strand at a region that flanks the target gene region that is to be measured. When each first primer for first cDNA strand synthesis is a gene-specific primer as discussed above, each second primer, which is also gene-specific such that each different first and second primer constitute a gene-specific primer pair, is preferably designed to anneal to a region located at a defined, preselected distance from the first primer binding site, so that the resultant double-stranded cDNAs produced by each gene-specific primer pair have the same or substantially the same lengths. Preferably, the gene-specific primer pairs are designed to produce cDNA products having lengths from about 50 to about 100 basepairs in length, and preferably from about 60 to about 70 basepairs in length.

While the preceding discussion describes the preparation of double-stranded DNA from mRNA templates, similar considerations apply for obtaining double-stranded DNA from other sources. For example, eukaryotic genomic DNA is typically obtained in double-stranded form, which is suitable for amplification. For single-stranded DNA species, double-stranded forms can be prepared by annealing and extending sequence-selective primers to the target DNA species, to form DNA duplexes containing complementary second strands. Also, the amplification of each target sequences can be accomplished using any number of primers and/or primer sets. For example, target amplification can be accomplished using a primer pair for each different

target sequence, by standard PCR methods, such as RT-PCR (reverse transcripase-PCR) for mRNA targets, as mentioned above. Alternatively, amplification can involve using a first primer to form a first complementary strand of a target, as discussed above, followed by amplification using a primer pair comprising second and third primers wherein the first primer is binds to a sequence in the target polynucleotide that is outside the target sequence flanked by the second and third primers. In another embodiment, amplification is by nested PCR, using a first primer set comprising first and second primers for one or more initial amplification cycles, followed by amplification using a second primer set comprising third and fourth primers which amplify a sub-sequence within the sequence amplified by the first primer set. Additional nested primer sets can also be used if desired.

Any amount of sample may be used to obtain double-stranded DNA, provided that the amount is sufficient to ensure the presence of all target polynucleotides, particularly low-abundance species, and provided that the relative levels of the target polynucleotides are statistically reliable. For example, for a target mRNA species that is expressed with a copy number of 1000 per cell, one cell may suffice to provide a sufficient amount of mRNA, to prepare double-stranded cDNA. On the other hand, for mRNAs that are present at levels of less than one copy per cell, tens, hundreds or thousands of cells may be necessary to provide an adequate representation of the actual levels in the sample.

With reference to Figure 1A, a sample containing a double-stranded DNA mixture is reacted with a plurality of sequence-selective primer pairs to form double-stranded copies of the defined regions in direct proportion to the target levels originally present in the double-stranded starting material. Any number of different sequence-specific primers may be used for measuring a desired number of target polynucleotides, depending on the needs of the user. Preferably, the number of primer pairs is between 2 and 500, more preferably between 5 and 300, and more preferably between 5 and 200. It is also preferred that the sequence-selective primers have similar melting temperatures (Tm values) to ensure similar binding kinetics with respect to their target sequences and to reduce non-specific priming. The Tm of each primer will depend on primer length, GC content, sequence, and on the characteristics of the surrounding solution (e.g., pH and salt concentrations). Preferably, the primers are designed to have Tm values that are within a 10°C range, and preferably within a 5°C range, under a particular pH and

salt concentration (Breslauer, cited below). Methods for estimating Tm values of primers have been described by Rychlik *et al.* (1989) Nucleic Acids Research 17:8543-8551 and (1990)18:6409-6412; Breslauer *et al.* (1986) Proc. Natl. Acad. Sci., 83: 3746-3750; and Wetmur (1991) Crit. Rev. Biochem. Mol. Biol., 26: 227-259, for example. Typically, primer lengths are from 10 to 40 nucleotides in length, more preferably from 22 to 26 nucleotides in length, although longer and shorter primers are also contemplated.

In a preferred embodiment, the defined target regions are amplified simultaneously by polymerase chain reaction (multiplex PCR) using sequence-selective primer pairs to generate larger amounts of the target segments for subsequent ligation and cleavage steps. Each primer pair is designed to flank a different, defined region of a target polynucleotide of interest, and each defined region contains a selected restriction endonuclease site. Preferably, the defined regions contain a restriction site for the same endonuclease. The defined target regions are selected to have lengths sufficient to uniquely identify the corresponding target polynucleotides, especially after the first cleavage step described below. For example, a target sequence that is 8 to 10 nucleotides in length can be sufficient to uniquely identify a target polynucleotide. There are 262,144 unique polynucleotides for a target sequence of 9 nucleotides (4^9). It is estimated that the human genome contains approximately 50,000 to 100,000 genes. Thus, a target sequence of 9 nucleotides can be sufficient to uniquely identify every expressed human gene. Also, if a target sequence of interest is derived from a non-human source, the target sequence length can be smaller.

The selected restriction endonuclease site (also referred to herein as recognition sequence) in each defined target region can be located anywhere within the defined target region, but it is preferably located at or near the middle of all different defined target regions, to facilitate production of cleaved fragments having similar lengths. Additional possible features of the restriction sites are discussed below.

Polymerase chain reaction cycles of denaturation, primer annealing, and primer extension are performed iteratively until the desired amounts of the target products are produced. The primers are added to the double-stranded DNA mixture, followed by denaturation of the DNA mixture to single-stranded form by standard methods, typically by applying high temperature of at least 90-95°C for an appropriate time. The mixture is allowed to cool to promote sequence-specific annealing of the primers opposite ends of

each target, followed by primer-extension mediated by an appropriate DNA polymerase. Preferably, the reaction mixture is subjected to at least one cycle, and preferably multiple cycles of denaturation, primer annealing, and primer extension, until the desired quantity of amplified product has formed. Preferably, the number of cycles is from 5 to 40, and typically is from 20 to 40.

According to an important aspect of the present invention, amplification should be performed under conditions effective to provide amplification yields that are generally uniform and consistent for all target sequences. In other words, the amplification conditions should not greatly favor amplification of any particular target sequence over any other target sequences. Otherwise, the relative quantities of different amplification products may not accurately reflect the original expression ratios in the sample.

To facilitate uniformity of target amplification, primers with similar or identical melting temperatures are preferably employed, as discussed above. Also, it is preferred that the gene-specific primer pairs be designed to produce amplification products having the same or substantially the same lengths. For example, the primer pairs may be designed to produce amplification products (copies) with lengths of from 60 to 70 base pairs, including the primer sequences and the defined target region that is amplified, and the lengths of the products are preferably within 10, and more preferably within 5 base pairs of each other. It is also recommended that the primer sequences be selected to minimize internal secondary structure. Shorter total amplification lengths are preferred over longer amplification lengths, to reduce levels of incomplete primer extension products and to increase the uniformity of amplification for the different targets.

Conditions for performing simultaneous (multiplex) PCR using multiple sequence-specific primer pairs are well known in the art (e.g., see Henegariu, O., et al., BioTechniques 23:504-511 (1997); Willey, J.C., et al., Am. J. Respir. Cell Mol. Biol. 17:114-124 (1997); and Apostolakos, M.J., et al., Anal. Biochem. 213:277-284 (1997). One advantage of co-amplification (simultaneous measurement in the same sample) is that each target provides a reference for comparison with the levels of other target in the sample. In an alternative approach, the sample targets may be amplified separately by non-multiplex competitive RT-PCR, wherein each target is amplified separately in the presence of a competitive template as an internal standard (e.g., see Gilliand, G., et al., Proc. Natl. Acad. Sci. USA 87:2725-2729 (1990); DeMuth, J.P., Am. J. Respir. Cell

Mol. Biol. 19:25-29 (1998). This method is advantageous for use with an air themocycler (e.g., from Idaho Technologies).

Typically, both primers for each primer pair are pre-loaded in each of the respective detection chambers, along with the standard nucleotide triphosphates (e.g., ATP, CTP, GTP, and TTP), or analogs thereof, for primer extension, and any other appropriate reagents, such as $MgCl_2$ or $MnCl_2$. A thermally stable DNA polymerase, such as Taq, Vent, or the like, may also be pre-loaded in the chambers, or may be mixed with the sample prior to sample loading.

For double-stranded DNA mixtures that are prepared from single-stranded targets, such as mRNAs, amplification can be combined with first and second cDNA strand synthesis by utilizing the same PCR primer pairs throughout cDNA synthesis and amplification, and by performing the second strand synthesis as the first cycle of PCR amplification. For example, in Example 3, an mRNA sample mixture is contacted with a plurality of different gene-specific primer pairs, each designed to form a defined-length cDNA product containing a defined region from a selected target gene sequence. The first primer in each pair anneals to the 3'-end of the target mRNA gene sequence that is to be copied, while the second primer in each pair remains as a bystander in solution. The annealed primer is extended with reverse transcriptase. The resultant cDNA/mRNA hybrid is denatured by appropriate conditions, and the second primer of each primer pair is annealed to the cDNA strands, while the first primer remains in solution as a bystander or anneals again to the complementary mRNA strand. The second primer is extended using a DNA polymerase in the presence of dNTPs and any other necessary reagent, to produce the desired double-stranded cDNA product. The steps of duplex denaturation, primer annealing, and primer extension are repeated a desired number of times by multiplex PCR using the same primer pairs to produce the desired amount of product.

In another embodiment, the gene-specific target segments of desired length are not formed until amplification by polymerase chain reaction, in which second pairs of gene-selective primers are employed for the first time for target DNA amplification. For example, if poly-T primers are used for first cDNA strand synthesis (all mRNAs are reverse-transcribed), and a first gene-specific primer is used for second cDNA-strand synthesis (forming cDNA strands that begin at their 5'-prime end with the gene-specific primer and potentially terminate at their 3'-ends with poly-T), it is necessary to amplify

both strands of the double-stranded cDNA products at least once to form shorter target segments having the desired lengths and termini. Exponential amplification in the presence of gene-specific primer pairs can then be performed to enrich for the desired length products relative to the first cDNA strands formed.

The gene-selective primer pairs are preferably designed to flank target regions that contain a single endonuclease restriction site of the same type, and preferably for a Type I endonuclease. Type I endonucleases recognize, bind to, and symmetrically cleave palindromic sequences in double stranded DNA. Site-specific cleavage by the Type I endonuclease yields cleavage products having symmetrical terminal sequences which may be blunt ended (no overhangs) or sticky ended (having single-strand extensions of one or more nucleotides in length). Preferably, the endonuclease yields fragments having sticky ends with overhangs of at least two, preferably at least three, and more preferably at least four nucleotides in length.

Since it is desired that each target gene contain a single restriction site for the same endonuclease, the restriction site should be one that occurs relatively frequently. For example, a recognition site consisting of four specifically required base pairs is expected to occur randomly with a frequency of once in every 256 bases (4^4), and would be expected (on average) to occur four times in a polynucleotide having a length of 1000 nucleotides. Endonucleases with tetrameric recognition sequences that produce fragments with sticky ends include DpnII, MboI, NlaIII, and Sau3A1, for example. Similarly, endonucleases are also known that recognize a split recognition sequence consisting of required base residues which are separated by one or more base residues of variable base type. An example of this type of endonuclease is DdeI, which recognizes a specific base sequence, CTNAG, which consists of four required residues (dimers CT and AG) separated by a variable nucleotide N which can be any of the four possible bases. Since DdeI cleaves between the C and T residues, DdeI produces cleavage fragments having 5'-overhangs that are three nucleotides in length. It will be appreciated that since the recognition sequence of DdeI requires four specific base residues, it is expected to have the same random frequency as endonucleases that recognize contiguous tetrameric sequences. Other endonucleases with split recognition sequences include BsaJI, Fau4HI, HinfI, and Sau96I, for example. Many suitable endonucleases are known in the art and are readily available from commercial sources such as New England Biolabs, Boehringer Mannheim, and Amersham Pharmacia

Biotech. Since the sequences of the target polynucleotides to be measured will be known beforehand, it is straightforward to scan all candidate target sequences for the locations of all known restriction sites and to choose one or more restriction sites that are common to all. Primer pairs may then be designed that flank the chosen restriction site and which are designed to produce double stranded products which are preferably of substantially the same lengths.

Treatment of the amplified DNA mixture with the selected endonuclease (Fig. 1A) generates a cleavage mixture of DNA fragments each having a sticky end produced by the endonuclease and a blunt end. The lengths of the cleaved fragments are preferably about the same. In Example 4, cleavage is accomplished using the Type I endonuclease NlaIII, which produces sticky ends terminating with 5'CATG-3'. The cleavage reaction divides each amplified target region into to two fragments, both of which may be suitable for ligation to adaptors as described further below.

From the cleavage mixture, at least two (first and second) aliquots are prepared for ligation to at least two different-sequence complementary adaptors. The first and second adaptors each contain (i) a first terminal end that is cohesive with the cleaved termini in the DNA fragments, (ii) a second terminal end that is non-cohesive with respect to the DNA fragments, so that only one end of each adaptor binds to a DNA fragment, and (iii) a unique-sequence segment that allows the first and second adaptors to be distinguished from each other, wherein the resultant conjugates comprise an adaptor segment and a target-identifier segment, by ligation of cohesive ends,

An exemplary structure for an adaptor that may be used in the invention is shown in Fig. 2. As shown, adaptor 40 contains a double-stranded region 42 which is flanked at one end by a defined sequence sticky end comprising a single-stranded region 44, which is complementary to the sticky ends of the target gene fragments in the cleavage mixture. Adaptor 40 also contains a distal end 46 that is designed to be substantially resistant to self ligation when a ligase enzyme is present, i.e., so that the distal end of one adaptor 40 is substantially incapable of enzyme-catalyzed ligation to the distal end of any other adaptor. For example, distal end 46 can comprise a 5' or 3' overhang that is not significantly complementary with itself, or it may include parallel 5'- and 3'-polynucleotides that are not complementary to one another, as with Adaptors #1 and #2 described in Example 4 (Fig. 3). In yet another approach, 3'- and 5'-terminal hydroxyl groups at distal end 46 may be capped with blocking groups, such as trityl, t-

Boc, or the like. The adaptor may also include a Type IIs restriction endonuclease recognition site, such as site 48, for use in an optional trimming step discussed further below.

The method of the present invention utilizes at least two different-sequence adaptors to prepare adaptor-fragment conjugates for subsequent ligation. Each different-sequence adaptor contains a unique-sequence segment 50 comprised of a polynucleotide sequence that is different for each different-sequence adaptor. The nucleotide sequences of the adaptors should be selected so as not to hybridize substantially with the target sequences or with other, different-sequence adaptors.

Each adaptor should have a length and sequence sufficient (1) to allow hybridization of an adaptor-specific primer to the adaptor, for polymerase chain reaction, (2) to distinguish different-sequence adaptors from each other, such that cross-hybridization between different-sequence adaptors is substantially inhibited, and (3) to encode a Type II restriction site if desired. The lengths of the adaptors are not critical, although it is preferred that the lengths are substantially the same. Preferably, the strands of each adaptor have lengths of from 20 to 60 nucleotides (shortest strand), more preferably from 20 to 45 nucleotides, and more preferably 35 to 45 nucleotides, although longer and shorter adaptors are also contemplated.

Returning to Figs. 1A-1B, each of the two or more aliquots from the endonuclease cleavage mixture is contacted with a selected different-sequence adaptor A_i under conditions effective to hybridize and ligate the adaptor to a complementary cleavage fragment, to form adaptor-fragment conjugates. In other words, a first adaptor A1 is reacted with a first aliquot to form conjugates with A1; a second adaptor A2 is reacted with a second aliquot to form conjugates with A2; and so on. Since each adaptor contains a sticky end that is complementary to the cleaved gene fragments, random ligation produces adaptor-fragment conjugates in direct proportion to the original levels of the source target polynucleotides in the sample. Typically, a 10:1 molar excess of each adaptor relative to total cleavage fragments is used to enhance the yield of the desired adaptor-fragment conjugates relative to non-productive fragment-fragment conjugates.

The conjugates that are produced by each ligation reaction contain an adaptor segment from adaptor 40 and a target-identifier segment from cleavage fragment 52, due to ligation of complementary cohesive ends. With reference to Fig. 2, adaptor segment

62 contains an external terminus 62a, also referred to as the distal end, which also defines a terminal end of the conjugate, and an internal terminal end 62b which is linked to the target-identifier segment 64. Each target-identifier segment 64 likewise contains an external terminal end 64a, which is blunt-ended as a result of PCR amplification (if the primers were perfectly complementary to the amplified sequences), and an internal end 64b, which is linked to the internal terminal end 62b of the adaptor segment.

Since cleavage of the amplified target regions produces two target-identifier segments for each amplified target polynucleotide, either or both of the segments can be used to quantify the source target polynucleotide, and it is expected that the two segments for a particular target polynucleotide should appear with the same frequency. Alternatively, adaptor ligation can be limited to one segment per target polynucleotide by using an endonuclease that produces complementary sticky ends having non-identical sequences, and by using adaptors that are complementary to only one of the sticky ends.

Optionally, the adaptors can include a Type IIs restriction site, such as recognition site 48 in adaptor 40, for trimming the ligated target-identifier segment to remove nucleotides to simplify analysis. Recognition site 48 is oriented in the adaptor so that the cleavage site 68 of the Type IIs endonuclease is located within the ligated target fragment at a selected distance from the junction between the adaptor and target fragment. Treatment of the conjugate with the Type IIs endonuclease is effective to sever a fragment 70 containing distal end 64a, yielding a conjugate 80 containing a shortened target-identifier segment 64c. A variety of Type IIs restriction enzymes are known, such as Bbs I, BbvI, BsmFI, and Fok I, for example, which have cleavage sites located at specific distances from their recognition sequences to produce cleavage products having blunt or stick ends. If the products have sticky ends, the sticky ends can be filled in with a DNA polymerase (e.g., Klenow fragment) in the presence of dNTPs to yield blunt ended products. Products from treatment with an endonuclease that generates blunt ends can also be subjected to a fill-in reaction with DNA polymerase to fill-in spurious (unexpected) sticky ends created during endonuclease treatment. Preferably, the target-identifier segments remaining in the conjugates have a length in the range of 10 to 50 base pairs, and more preferably from 10 to 20 base pairs. In Example 4, the adaptors contain a BsmFI site at the sticky ends of the adaptors. Treatment of the adaptor conjugates with BsmFI cleaves the target-identifier segments

at a linkage 14 base pairs away from the BsmFI recognition site, so that the resultant identifier segments are exactly 14 base pairs long.

The conjugates may be purified to enrich for the desired adaptor-fragment conjugates, e.g., by gel electrophoresis or column chromatography. The desired adaptor-fragment conjugates can be easily separated from other species on the basis of size, particularly if the fragments and adaptors have different lengths, e.g., differing by at least two, preferably by at least five nucleotides in length, so that the chromatographic or electrophoretic properties of the adaptor-adaptor and fragment-fragment conjugates differ significantly from the chromatographic or electrophoretic properties of the adaptor-fragment conjugates.

Following conjugate formation, the aliquots are recombined and are reacted under conditions effective to promote ligation only between external termini of the target-identifier segment of the conjugates, to form conjugate dimers. With reference to Figs. 1A-1B, the adaptor-fragment conjugates are blunt-end ligated using an appropriate DNA ligase to produce conjugate dimers that contain two internal, abutting target-identifier segments, and two external, flanking adaptor segments. Assuming that only two different conjugation reactions have been performed (i.e., using two different adaptors as discussed above), blunt-end ligation of the conjugates produces a mixture of dimer products in an approximately 1:2:1 ratio ([A1/A1 dimer]:[A1/A2 dimer]:[A2/A2 dimer]). In the dimers produced as described in Example 4, the final lengths of the desired conjugate dimers are about 102 to 108 bp. In Fig. 1B, the following ligation products are shown: A1-S1-S2-A1, A1-S5-S7-A2, and A2-S11-S4-A2, where A1 and A2 represent the two adaptors, and each Sn represents a different, random target-identifier fragment. More generally, the mixture contains species of the form An-Si-Sj-Am, where An and Am represent adaptors A1 and A2 and may be the same or different, and Si and Sj represent all possible combinations of the target-identifier fragments.

The dimers are amplified using first and second primers for annealing to the unique-sequence segments of the adaptors. The primers are added to the dimer solution, and amplification is performed as above, until the desired quantity of amplified product has formed. Preferably, the number of cycles is from 5 to 40, and typically is from 15 to 35. During amplification, only dimers containing an adaptor A1 and an adaptor A2 are amplified (A1-Si-Sj-A2), because dimers that contain only one adaptor type (e.g., two copies of adaptor A1) readily form pan-handle structure 90 as illustrated in Fig. 4.

Unimolecular self-annealing occurs in these strands much more rapidly than annealing to primers because the complementary A1 termini at each end of the individual dimer strands are maintained in close proximity to each other, so that the likelihood of self hybridization is very high. The formation of a pan-handle structure effectively prevents amplification of an A1/A1 or A2/A2 strand during each primer-extension step. Dimers containing an A1 adaptor and an A2 adaptor, however, can be successfully amplified because the A1 and A2 terminal sequences are sufficiently different to prevent hybridization with each other.

After the conjugate dimers have been amplified, the adaptor segments can be removed by treatment with a selected restriction enzyme that cleaves the junction between the adaptors and the target-identifier segments without cleaving the junction between the two target-identifier segments. Conveniently, the restriction endonuclease is the same Type I endonuclease that was used prior to ligating the sample fragments to the adaptors. The resultant target-identifier dimers (TID dimers) of the form Si-Sj (Fig. 1B) are then isolated from the cleaved adaptor fragments by any suitable technique. For example, TID dimers may be separated on the basis of size, e.g., by polyacrylamide gel electrophoresis or chromatography. Size-based separations can be facilitated by designing the adaptors to have lengths that are significantly longer than the TID dimers. In the protocol described in Example 4, endonuclease cleavage of the conjugate dimers yields TID dimers having lengths of about 41 and 39 basepairs if primers corresponding to the single-underlined (thin underline) sequences in Adaptors #1 and #2 in Example 4 (Fig. 3) are used (GGATTTGCTGGTGCAGTACA, residues 4-23 of SEQ ID NO:8, and CTGCTCGAATTCAAGCTTCT, residues 4-23 of SEQ ID NO:10), or about 18 basepairs if a second set of primers corresponding to the thickly underlined sequences is used (AGGCTTAATAGGGACAT, residues 27 to 43 of SEQ ID NO:8, and ACGATGTACGGGGACAT, residues 25 to 41 of SEQ ID NO:10). Note that if the target-identifier is derived from a gene, then the target-identifier dimers can also be referred to as gene-identifier dimers.

Following release of the adaptors, the TID dimers may be polymerized into multimers using appropriate ligation conditions to join the TID dimer segments end-to-end. The TID dimer units are hybridized to each other via their complementary cohesive ends and are ligated with an appropriate ligating agent. The multimers may be of any desired length, and are preferably at least 300 nucleotide bases in length, and are

typically between 20 and 2000 bases in length, and more typically between 400 and 800. The average length and length distribution of the multimers will depend on TID abundance, the choice of ligation agent, temperature, and the like, according to standard principles of polymerization. Polymerization parameters can be optimized readily by varying conditions and examining the resultant length distributions by polyacrylamide gel electrophoresis or other methods.

Conveniently, the polymerization conditions may include a terminating adaptor for capping the ends of multimers to further control multimer lengths, and to introduce amplifiable primer segments to the ends of the multimers as shown in Figs. 1A-1B. The terminating adaptor is designed to be substantially non-hybridizable with the target-identifier segments as well as any other potential TID multimer sequences that might arise due to polymerization. Preferably, the Tm of the primer for its complementary sequence in the terminator adaptor sequence is designed to have a Tm that is at least 5°C, and preferably at least 10°C, greater than the Tm of the primer for binding any other potential multimer sequence. In Example 4, polymerization of TIDs was performed in the presence of a double-stranded terminator adaptor designated A3. The multimers are amplified until the desired level of multimers is achieved. In Example 4, multimers are amplified using a PCR primer, designated PDS1-CUA, which contains a sequence complementary to one of the strands of the A3 adaptor, for amplification of both strands of each multimer. Typically, 20 to 25 PCR cycles are adequate. Further guidance on amplification of multimers using a single primer can be found in PCT Publication No. WO 89/12695, and Morgan, J.G., *et al.* (1992) Nucleic Acids Res. 20: 5173-5179.

After polymerization, the relative abundances of target-identifier segments are quantified by counting the frequencies of the identifiers in one or more multimers, to provide an estimate of the levels of the target polynucleotides that were originally present in the sample. This analysis may be performed using any suitable method selected by the user. Typically, the requisite data are obtained by sequence analysis of one or more different multimers by any known sequencing method. The contiguous identifier segments can be readily identified from their sequences without needing intervening spacer segments, since the sequences of the identifier segments are known beforehand. For example, the product of polymerization in Fig. 1 contains target-identifier segments S1-S17-S31-S17-S10, such that the segment S17 occurs twice. In

theory, if a small number of target polynucleotide sequences is amplified and if the multimer is sufficiently long, sequencing of a single multimer can provide a sufficient amount of data to determine target levels.

In another alternative approach, the multimers can be immobilized onto a solid support, e.g., using biotin-avidin pairing, and the various identifiers can be quantified by hybridizing labeled probes that are complementary to each identifier segment.

In another approach, the identifiers can be quantified by mass spectrometry. For example, the molecular masses of multimers containing up to 2000 nucleotides or more can be readily determined by mild techniques such as matrix-assisted laser desorption/ionization (MALDI) (see, for example, Berkenkamp et al., Science 281:260-262 (1998); Alper, Science 279:2044 (1998); Hung et al., Anal. Chem. 70:3088 (1998); and Monforte et al., Nature Med. 3:360 (1997). Since the molecular masses of large polynucleotides can be determined with high accuracy, and since each of the standard nucleotide bases has a distinct molecular mass, the TID composition of each multimer can be determined by comparing the observed TID mass against linear combinations of the masses of the potential constituent TIDs, provided that each TID has a distinct base composition. Multimers can also be sequenced as nested fragments, using, for example, the mass spectrometry method described by K$^{\text{Ⓟ}}$ster (WO 94/16101).

To facilitate sequencing, the multimer products can be incorporated into a suitable vector to form a library of multimers for cloning and sequence analysis. For this purpose, the terminating adaptor preferably includes a restriction endonuclease site that corresponds to an insertion site in the vector. Preferably, the endonuclease site does not occur in the multimers. This criterion can be met by preparing a list of possible multimer sequences that might result from polymerization (via random ligations) of target-identifier sequences from the target genes, and comparing them with candidate restriction sites for the terminator adaptors. For this purpose, longer restriction sites are preferable (e.g., comprising at least six specific base positions), since the frequency of restriction sites decreases with longer recognition sequences. Cleavage of the capped multimers with the selected restriction endonuclease produce multimer products with the desired cohesive ends for insertion into the vector. Cohesive ends can also be introduced at the termini of multimers using PCR primers that have protruding 5'-ends, to create 5'-overhangs for ligation. Multimers can also be inserted into a selected vector by blunt end ligation. The multimer mixture is treated with the appropriate endonuclease

and ligated to the vector linearized with the same endonuclease or isoschizomer thereof. The recombinant vectors are transformed into a suitable host organism and are plated. Clones from the library can be selected randomly and sequenced until the desired data is obtained.

It is not necessary that all of the multimers be cloned into vectors, nor is it essential that particular multimers are cloned. Since the various TIDs are polymerized randomly, the necessary abundance information can be obtained by sequencing a certain number of TIDs in a random multimers, until the desired level of certainty is obtained for statistical purposes.

Since the expression of different genes can vary by several orders of magnitude, transcript levels for one or a few genes can sometimes dominate less abundant transcripts. This may not be a problem if the levels of the low abundance transcripts are not of interest. Otherwise, significantly greater amounts of data must be collected to adequately sample low abundance species to satisfy statistical requirements. Therefore, to reduce the amount of data collection, gene-specific primer pairs can be grouped into selected sets according to known or expected levels of their target polynucleotides. For example, a first primer set can be designed to co-amplify targets having abundances on the order of about 1000 copies per cell to about 10,000 copies per cell, another set for targets in a range of about 100 copies per cell to about 1000 copies per cell, another for a range of about 10 copies per cell to about 100 copies per cell, and another set for abundances of about 10 copies or less. In another example, three sets are used which cover ranges of about 5000 or greater, about 100 copies per cell to about 5,000 copies per cell, and about 100 copies per cell or less. The particular range(s) will be selected according to the preferences of the user.

The reliability of measurements can be evaluated using multiple aliquots from the same sample and computing an average and standard deviation. Typically, the measured levels are highly reproducible for a particular sample. Absolute levels of target polynucleotides can be calculated by adding control polynucleotides to the sample before the samples are processed. Additional reliability is provided by comparative profiling of target levels between different samples, such that targets whose relative levels appear to be constant for all samples can be used to normalize the data, to highlight targets whose levels are significantly different between samples.

It will be appreciated that the accuracy of the measured levels is dependent on the quantity of multimers that are sequenced, and the number and length of the target-identifier segments. In general, the precision and accuracy of measured expression levels increases as the number of sequenced multimers is increased, subject to approaching a point of diminishing returns. Generally, counting will be sufficient once the determined relative amounts remain substantially the same as additional counting data is incorporated, or once the desired qualitative information has been obtained. Statistical methods may also be applied as appropriate, in accordance with well known mathematical methods.

Differential target levels from different samples can be detected from variations in the counts of cognate target sequence identifiers. A preferred method for characterizing abundance data is the rigorous significance test developed by S. Audic and JM Claverie, (1997) Genome Res. 1997:986-995. A computer program by these authors that allows computation of the confidence intervals corresponding to arbitrary significance levels and sampling size is currently available on the world wide web at http://igs-server.cnrs-mrs.fr)

Utility. The present invention provides methods that are useful for simultaneously measuring large numbers of selected target polynucleotides in a sample. The invention is particularly advantageous for quantifying or monitoring target polynucleotides that have low abundances, e.g., mRNA transcripts that are present on average at less than one copy per cell, since the target genes can be amplified to any desired level without interference from more abundant species. The invention is thus useful for comparing expression profiles of a sample, such as a cell or tissue sample, in a first state versus a second state.

The present invention finds a number of useful applications, including:

1. Development of comprehensive cell or tissue abundance profiles for selected targets. For example, this provides a convenient way to obtain expression profiles of selected target genes in different cell-types and tissue-types, to establish reference levels and evaluate how expression profiles vary.

2. Determination of target polynucleotide levels under various conditions, such as normal conditions, different stages of development, infection by a pathogen, proliferative states (e.g., cancer), diseased states, etc. The invention can be used to establish how the

levels of the target polynucleotides vary among different samples or the same sample-types under different conditions.

3. Screening for genes needed to maintain health, or which may underlie disease.

4. Discovery and evaluation of diagnostic markers. The invention can be used to identify target polynucleotides with elevated or suppressed levels relative to a defined control.

5. Discovery and evaluation of known or candidate therapeutic agents *in vivo* or *in vitro*. The invention can be used to determine whether and how levels of selected target polynucleotides change in response to the addition of therapeutic agents or to non-chemical therapeutic treatments. The invention is useful for evaluating whether agents are efficacious and what dosage levels are effective. The measured levels can also be used to investigate the mode of action of chemical agents, and the effects of polymorphisms and mutations.

6. Identification of targets for gene therapy.

7. Characterization of effects of infection on infected cells or tissues.

8. Investigation of changes in target levels in response to environmental changes or during specific biological processes.

For example, the differential expression of genes in different tissues, during development or during specific pathologies, is of foremost interest to both basic and pharmaceutical research, especially in the absence of functional clues regarding the genes of interest. Evidence of differential expression is often the most important criterion to prioritize the exploitation of anonymous sequence data in both basic and corporate research. For example, decreased levels may result from damage to transcription regulation sequences, increased expression of a suppressor, or a decrease in activator. The present invention provides a powerful method to reveal insights into gene function and gene pathways, and to focus on new drug targets. In contrast to functional assays, the quantitative analysis of gene expression level lends itself to large-scale implementation.

Furthermore, the present invention provides a way to focus on relatively small groups of disease candidate genes, which may have been identified by genetic linkage analysis, EST analysis, or in any other way, to allow rapid characterization of their expression levels in selected tissues and conditions. By focusing on selected target genes, the method avoids problems associate with processing vast amounts of redundant or extraneous data which has plagued previous methods.

Fig. 5 shows a bar plot of levels of selected gene sequences from mRNA transcripts obtained from CD4+ cells from bronchoalveolar lavage (BAL) samples from normal and asthmatic subjects using the methods described in Examples 1 through 4. The subjects included three healthy subjects (bars 1-3 on the left of each group) and four asthmatic subjects (bars 4-7, on the right of each group). As can be seen, the levels of nearly all target gene transcripts were relatively constant among healthy subjects, asthmatic subjects, and both groups taken together. However, the target designated P2 showed marked variability in expression levels that spanned an 8-fold range. In one asthmatic subject, the level of this target was twice as great as the highest level measured in the normal subjects. mRNA for IL10, which is known to be associated with asthma (e.g., Robinson, D.S., et al., Am. J. Respir. Cell Mol. Biol. 14:113-117, 1996) showed a similar pattern, except that the absolute levels were several-fold lower than those observed for P2. A similar correlation between IL10 and P2 is observed for blood samples, as shown in Fig. 6.

The results demonstrate how the invention can be used to identify polynucleotides whose levels are altered in a disease state, and which may cause or result from the presence of a disease or other abnormal state. The data also confirm the reproducibility of the method, given the relatively constant levels that were measured for several of the targets. Similar approaches can be used for other sample types, such as tumor cells, to characterize expression profiles, develop new diagnostic methods, and test the effects of therapeutic methods on gene expression.

The following section describes another aspect of the invention, which concerns polynucleotides corresponding to the above-described P2 gene target sequence. These polynucleotides encode a novel polypeptide, designated herein as P2, the expression of which is significantly up-regulated in tissues of asthma patients as described above.

II. Novel Biomolecules Up-Regulated in Asthmatic Tissues

The present invention provides, in another aspect, polynucleotides and polypeptides that are useful for a variety of applications, particularly relating to asthma.

A. Definitions

A "CX5C chemokine" is defined herein as a small (under 100 amino acids), preferably basic (pI > 8) polypeptide containing at least four cysteines, the two N-

terminal cysteines separated by five amino acids (i.e., containing the sequence motif CXXXXXC, where X is any amino acid).

"P2" refers to a polypeptide which (i) is a member of the novel CX5C chemokine family as defined above and (ii) has at least 80 percent, preferably at least 85 percent, more preferably at least 90 percent, and still more preferably at least 95 percent amino acid sequence identity to SEQ ID NO:2. The polypeptide may be a mature P2 polypeptide and/or a modified P2 polypeptide.

The term "mature P2 polypeptide" refers to the P2 polypeptide as it exists after post-translational processing, *e.g.* removal of a signal sequence.

The term "modified", when referring to a polypeptide of the invention, means a polypeptide which is modified either by natural processes, such as alternative splicing or proteolytic processing or other post-transcriptional or post-translational modifications, or by chemical modification techniques which are well known in the art. Among the numerous known modifications which may be present include, but are not limited to, acetylation, acylation, amidation, carboxylation, ADP-ribosylation, glycosylation, GPI anchor formation, covalent attachment of a lipid or lipid derivative, methylation, myristoylation, pegylation, prenylation, phosphorylation, ubiqutination, or any similar process.

The term "biologically active" refers to a P2 having structural, regulatory or biochemical functions of a naturally occurring P2, such as the ability to interact with a P2-binding protein (*e.g.*, a P2-receptor). Likewise, "immunologically active" defines the capability of a natural, recombinant or synthetic P2 or any fragment thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "fragment," when referring to P2, means a polypeptide which has an amino acid sequence which is the same as part of but not all of the amino acid sequence of P2, which either (a) retains essentially the same biological function or activity as P2, or (b) retains at least one of the functions or activities of P2, or (c) which is capable of interacting with P2, or with a protein or other molecule which binds P2 (*e.g.*, a P2-receptor), to alter a function or activity or the cellular/subcellular localization of P2 or of a protein which interacts with P2. Exemplary fragments contemplated include: a fragment which inhibits or potentiates the binding of P2 to its receptor, said receptor which, in the absence of the fragment, binds P2 and either modulates the activity of P2

or modulates a cellular process in which P2 participates. Another fragment contemplated is one which retains immunological activity of P2. The fragment preferably includes at least 10 contiguous amino acid residues of P2, more preferably at least 15 or 20 residues.

A polypeptide containing a "portion" of a polypeptide of the invention means a polypeptide which contains an amino acid sequence which is the same as part of the amino acid sequence of the present invention or a variant thereof, and which does not necessarily retain any biological function or activity.

A "conservative substitution" refers to the substitution of an amino acid in one class by an amino acid in the same class, where a class is defined by common physicochemical amino acid sidechain properties and high substitution frequencies in homologous proteins found in nature (as determined, e.g., by a standard Dayhoff frequency exchange matrix or BLOSUM matrix). Six general classes of amino acid sidechains, categorized as described above, include: Class I (Cys); Class II (Ser, Thr, Pro, Ala, Gly); Class III (Asn, Asp, Gln, Glu); Class IV (His, Arg, Lys); Class V (Ile, Leu, Val, Met); and Class VI (Phe, Tyr, Trp). For example, substitution of an Asp for another class III residue such as Asn, Gln, or Glu, is a conservative substitution.

A "non-conservative substitution" refers to the substitution of an amino acid in one class with an amino acid from another class; for example, substitution of an Ala, a class II residue, with a class III residue such as Asp, Asn, Glu, or Gln.

"Optimal alignment" is defined as an alignment giving the highest percent identity score. Such alignment can be performed using a variety of commercially available sequence analysis programs, such as the local alignment program LALIGN using a ktup of 1, default parameters and the default PAM. A preferred alignment is the pairwise alignment using the CLUSTAL-W program in MACVECTOR, operated with default parameters, including an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM30 similarity matrix. If a gap needs to be inserted into a first sequence to optimally align it with a second sequence, the percent identity is calculated using only the residues that are paired with a corresponding amino acid residue (i.e., the calculation does not consider residues in the second sequences that are in the "gap" of the first sequence).

"Percent sequence identity", with respect to two amino acid or polynucleotide sequences, refers to the percentage of residues that are identical in the two sequences

when the sequences are optimally aligned. Thus, 80% amino acid sequence identity means that 80% of the amino acids in two optimally aligned polypeptide sequences are identical.

A first polypeptide region is said to "correspond" to a second polypeptide region when the regions are essentially co-extensive when the sequences containing the regions are aligned using a sequence alignment program, as above. Corresponding polypeptide regions typically contain a similar, if not identical, number of residues. It will be understood, however, that corresponding regions may contain insertions or deletions of residues with respect to one another, as well as some differences in their sequences.

The term "sequence identity" means nucleic acid or amino acid sequence identity in two or more aligned sequences, aligned as defined above.

"Sequence similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide. Thus, 80% amino acid sequence similarity means that 80% of the amino acid residues in two or more aligned polypeptide sequences are conserved amino acid residues, i.e. are conservative substitutions.

"Hybridization" includes any process by which a strand of a nucleic acid joins with a complementary nucleic acid strand through base-pairing. Thus, strictly speaking, the term refers to the ability of the complement of the target sequence to bind to the test sequence, or vice-versa.

"Hybridization conditions" are based on the melting temperature (Tm) of the nucleic acid binding complex or probe and are typically classified by degree of "stringency" of the conditions under which hybridization is measured. For example, "maximum stringency" typically occurs at about Tm-5°C (5° below the Tm of the probe); "high stringency" at about 5-10° below the Tm; "intermediate stringency" at about 10-20° below the Tm of the probe; and "low stringency" at about 20-25° below the Tm. Functionally, maximum stringency conditions may be used to identify nucleic acid sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify nucleic acid sequences having about 80% or more sequence identity with the probe.

An example of "high stringency" conditions includes hybridization at about 65°C in about 5x SSPE and washing at about 65°C in about 0.1x SSPE (where 1x SSPE = 0.15 sodium chloride, 0.010 M sodium phosphate, and 0.001 M disodium EDTA).

The term "gene" as used herein means the segment of DNA involved in producing a polypeptide chain; it may include regions preceding and following the coding region, e.g. 5' untranslated (5' UTR) or "leader" sequences and 3' UTR or "trailer" sequences, as well as intervening sequences (introns) between individual coding segments (exons).

An "isolated polynucleotide having a sequence which encodes P2" is a polynucleotide which contains a sequence encoding P2 (i) in isolation, and/or (ii) in combination with additional coding sequences, such as fusion protein or signal peptide, in which the P2 coding sequence is the dominant coding sequence, and/or (iii) in combination with non-coding sequences, such as introns and control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host, and/or (iv) in a vector or host environment in which the P2 coding sequence is a heterologous gene.

The terms "heterologous DNA" and "heterologous RNA" refer to polynucleotides that are not endogenous to the cell or part of the genome in which they are present. Generally such nucleotides have been added to the cell by transfection, microinjection, electroporation, or the like. Such polynucleotides generally include at least one coding sequence, but this coding sequence need not be expressed.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The term "fragment," when referring to a P2 polynucleotide sequence, means a polynucleotide which has a nucleic acid sequence which is the same as part of but not all of the nucleic acid sequence of the P2 gene or the P2 coding sequence. The polynucleotide fragment preferably includes at least 15 contiguous nucleotides of P2 coding sequence, more preferably at least 20 or 30 nucleotides.

The term "expression vector" refers to vectors that have the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and

eukaryotic expression vectors are commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

The term "substantially purified" refers to molecules, either polynucleotides or polypeptides, that are removed from their natural environment, isolated or separated, and are at least 60% free (by weight), preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "variant" polynucleotide sequence encodes a "variant" amino acid sequence which is altered by one or more amino acids from the reference polypeptide sequence. The variant polynucleotide sequence may encode a variant amino acid sequence which contains one or more "conservative" substitutions, wherein the substituted amino acid has structural or chemical properties similar to the amino acid which it replaces. In addition, or alternatively, the variant polynucleotide sequence may encode a variant amino acid sequence which contains "non-conservative" substitutions, wherein the substituted amino acid has dissimilar structural or chemical properties to the amino acid which it replaces. Variant polynucleotides may also encode variant amino acid sequences which contain amino acid insertions or deletions, or both.

An "allelic variant" is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

"Alternative splicing" is a process whereby multiple polypeptide isoforms are generated from a single gene, and involves the splicing together of nonconsecutive exons, and/or the incorporation of one or more intronic sequences, during the processing of some, but not all, transcripts of the gene. Thus a particular exon may be connected to any one of several alternative exons (and occasionally, introns) to form messenger RNAs. The alternatively-spliced mRNAs produce polypeptides ("splice variants" or "splice isoforms") in which some parts are common while other parts are different.

"Splice isoforms", or, alternatively, "splice variants" of P2, when referred to in the context of an mRNA transcript, are mRNAs produced by alternative splicing of the P2 gene. An exemplary splice isoform nucleic acid sequence is identified herein as SEQ ID NO:3.

"Splice isoforms", or, alternatively, "splice variants" of P2, when referred to in the context of the polypeptide itself, are P2 translation products which are encoded by

alternatively-spliced P2 mRNA transcripts. An exemplary P2 splice isoform polypeptide is identified herein as SEQ ID NO:4.

A "mutant" amino acid or polynucleotide sequence is a variant amino acid sequence, or a variant polynucleotide sequence which encodes a variant amino acid sequence, which has significantly altered biological activity from that of the naturally occurring polypeptide.

A "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

An "insertion" or "addition" is a change in a nucleotide or amino acid sequence which results in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring sequence.

A "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.

The term "modulate" as used herein refers to the change in activity of the polypeptide of the invention. Modulation may relate to an increase or a decrease in biological activity, binding characteristic, or any other biological, functional, or immunological property of the molecule.

The term "agonist" as used herein, refers to a molecule which modulates the activity of P2 by inducing, increasing, potentiating, or prolonging the duration of, a biological activity mediated by P2. Agonists may themselves be polypeptides, nucleic acids, carbohydrates, lipids, or derivatives thereof, or any other molecules which bind to and modulate the activity of P2.

The term "antagonist" as used herein, refers to a molecule which modulates the activity of P2 by blocking, decreasing, inhibiting, or shortening the duration of, a biological activity mediated by P2. Antagonists may themselves be polypeptides, nucleic acids, carbohydrates, lipids, or derivatives thereof, or any other molecules which bind to and modulate the activity of P2.

The term "humanized antibody" refers to an antibody molecule in which one or more amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding activity of the antibody.

"Treating a disease" refers to administering a substance effective to reduce the symptoms and/or severity of the disease.

B. Polynucleotides Encoding P2

The invention provides isolated polypeptides, identified herein as P2 and P2 splice isomers, and isolated polynucleotides which encode the polypeptides. As defined more fully in Section C below, P2 comprises a polypeptide (i) belonging to the novel CX5C chemokine family and (ii) having at least 80%, preferably at least 85%, more preferably at least 90%, or 95% sequence identity to the amino acid sequence identified as SEQ ID NO:2.

B1. P2 Gene and Transcript Sequences

The P2 gene is located on human chromosome 5 within the cytokine gene cluster at 5q31. The long arm of chromosome 5 is the site of a dense cluster of immunomodulatory genes encoding cytokines, cytokine receptors and transcription regulation factors. This region contains genes encoding IL3, GM-CSF, Il4, IL5, IL13, IL9, FGFA, IL12 (p40), HBEGF, IRF1, TCF1, EGR1, GRL1, CD14, and others. The human P2 gene is located within this cytokine gene cluster [upstream of IL13 and IL4.

Human genomic DNA corresponding to a P2 gene of the present invention is identified herein as SEQ ID NO:5. Northern blot hybridization experiments using MASTER BLOTS (Clontech) probed with a human P2 genomic probe showed that the most abundant expression among the tissues tested is in adult and fetal liver, where three potential P2-encoding transcripts were detected, of approximately 0.8 to 1.0 kb, 1.8 to 2.0 kb, and 4.4 kb, in order of decreasing relative abundance.

RT-PCR on 15 different human tissues revealed the presence of at least P2 splice isoforms, which are unequally represented in the tissues tested. In fetal liver and total embryo a shorter transcript was more abundant; in T cells, placenta, pituitary, testes, brain, spleen, lymph node and fetal lung, a longer transcript was more abundant; and in kidney and small intestine the two mRNA species were about equally represented. The corresponding mouse gene (discussed further below) is predominantly expressed as a single mRNA species of 0.8-0.9 kb in mouse embryo and in adult mouse liver. Human P2 gene activity may be regulated via alternative splicing, in contrast to the mouse gene.

The exemplary P2 genomic sequence contains three exons and two introns as follows:

Exon 1: nucleotides ~1047 to 1353 (or, alternatively, ~1105 to 1353) of SEQ ID NO:5. The translation ATG start codon is located at position 1297 of SEQ ID NO:5.

Intron 1: nucleotides 1354 to 1551 of SEQ ID NO:5. Intron 1 is incorporated into the P2B splice isoform coding sequence, as discussed below.

Exon 2: nucleotides 1552 to 1691 of SEQ ID NO:5.

Intron 2: nucleotides 1692 to 2011 of SEQ ID NO:5.

Exon 3: nucleotides 2012 to 2316 (or 2012 to 2633) of SEQ ID NO:5.

Initially synthesized P2 mRNA transcripts appear to start at nucleotide ~1047 or ~1105 of SEQ ID NO:5 due to the potential presence of dual promoters, and may terminate either at nucleotide 2316 or 2633 of SEQ ID NO:5, due to the presence of several polyadenylation signals within the 3'-UTRs.

From further analysis, sequences encoding human P2 and a splice isoform were determined. The polynucleotide sequence of a P2 transcript (also designated as splice isoform P2-A; SEQ ID NO:1) is 1459 nucleotides in length and consists of exon 1, exon 2, and exon 3, joined consecutively. The P2 open reading frame (ORF) extends from nucleotides 192 to 422 of SEQ ID NO:1, and encodes a 77 amino acid polypeptide identified as SEQ ID NO:2. A mouse P2 coding sequence, identified herein as SEQ ID NO:6, contains an open reading frame which encodes a 76 amino acid mouse P2 homolog, identified herein as SEQ ID NO:7.

The polynucleotide sequence of an exemplary human P2 splice isoform designated P2-B (SEQ ID NO:3) is 1657 nucleotides in length and consists of exon 1, intron 1, exon 2, and exon 3. The P2-B open reading frame extends from nucleotides 192 to 311 of SEQ ID NO:3 and encodes a 40 amino acid polypeptide identified as SEQ ID NO:4. The smaller P2-B translation product is due to the presence of an in-frame TGA stop codon within intron 1.

Other splice isoforms of P2 are contemplated by the present invention, with coding sequences comprising exons and introns as described above joined in consecutive or nonconsecutive order.

Of particular note is the presence of at least two ATTTA sequence motifs in the 3' untranslated regions (3' UTRs) of the P2 transcripts described above, which is a feature of mRNAs with short half-lives. Such "mRNA instability" sequence motifs are characteristic of the 3' UTRs of transiently-expressed mRNAs, such as those encoding

oncogenes, cytokines and chemokines (Akashi M, *et al.* (1994) Blood 83(11):3182-3187).

B2. Polynucleotide Compositions

The polynucleotides of the invention include those having sequences which encode P2 and P2 splice isoforms such as the embodiments described above, sequences complementary to the coding sequence, and polynucleotides which hybridize under at least high-stringency conditions to such sequences. The polynucleotides may be in the form of RNA or in the form of DNA, and include mRNA, cRNA, synthetic RNA and DNA, cDNA, genomic DNA, and antisense derivatives thereof. The DNA may be double-stranded or single-stranded, and if single-stranded may be the coding strand or the non-coding (i.e., complementary) strand.

In a general embodiment, the polynucleotide has at least 70%, preferably at least 80%, 90%, or 95% sequence identity with the P2 sequence identified as SEQ ID NO:1 or the splice isoform SEQ ID NO:3, or with the open reading frames (ORFs) corresponding to nucleotides 192-422 of SEQ ID NO:1, or nucleotides 192-311 of SEQ ID NO:3. In other embodiments, the polynucleotide has at least 70%, preferably at least 80%, 90%, or 95% sequence identity with one or more of the exons or introns as identified in Section B1 above, either singly or in combination, and, if in combination, joined together in consecutive or non-consecutive order. In other embodiments, the polynucleotide has a sequence (a) essentially identical to any of the sequences described above, or b) to its complement, or (c) hybridizes under at least high-stringency conditions to (a) or (b).

The polynucleotide may include a coding sequence of P2 or a splice isoform (i) in isolation, and/or (ii) in combination with additional coding sequences, such as fusion protein or signal peptide, in which the P2 coding sequence is the dominant coding sequence, and/or (iii) in combination with sequences such as introns, control elements, promoter elements, terminator elements, 5' and/or 3' untranslated regions, effective for, e.g., expression of the coding sequence in a suitable host, and/or (iv) in a vector or host environment in which the P2 coding sequence is a heterologous sequence.

The polynucleotide may encode a fragment of P2, corresponding to, for example, an immunogenic fragment.

The polynucleotides of the present invention may also have the polypeptide coding sequence fused in-frame to a marker sequence which allows for purification of P2 or its splice variants. The marker sequence may be, for example, a hexahistidine tag to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., *et al.* (1984) Cell 37:767).

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, which hybridize under high-stringency conditions to a P2 polynucleotide described above. These oligonucleotides have at least 15 bases, preferably at least 20 or 30 bases, and may be used as probes, primers, antisense agents, and the like, according to known methods.

B3. Preparation of Polynucleotides

The polynucleotides may be obtained by screening cDNA libraries, using oligonucleotide probes which can hybridize to or PCR-amplify polynucleotides which encode the P2 polypeptide and isoforms disclosed above. cDNA libraries prepared from a variety of tissues are commercially available (*e.g.*, from Clontech, Palo Alto, CA) and procedures for screening and isolating cDNA clones are well-known to those of skill in the art. Such techniques are described in, for example, Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Cold Spring Harbor Press, Plainview, N.Y., and Ausubel FM *et al.* (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

The polynucleotides may be extended to obtain upstream and downstream sequences such as promoters, regulatory elements, and 5' and 3' untranslated regions (UTRs). Extension of the available transcript sequence may be performed by numerous methods known to those of skill in the art, such as by PCR or primer extension (Sambrook *et al.*, supra), or by the RACE method using, for example, the MARATHON RACE kit (Clontech, Cat. # K1802-1).

Alternatively, the technique of "restriction-site" PCR (Gobinda *et al.* (1993) PCR Methods Applic. 2:318-22), which uses universal primers to retrieve flanking sequence adjacent a known locus, may be employed. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region.

The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia T *et al.* (1988) Nucleic Acids Res 16:8186). The primers may be designed using OLIGO(R) 4.06 Primer Analysis Software (1992; National Biosciences Inc, Plymouth, Minn.), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68-72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom M. *et al.* (1991) PCR Methods Applic 1:111-19) is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into a flanking part of the DNA molecule before PCR.

Another method which may be used to retrieve flanking sequences is that of Parker, JD *et al.* (1991; Nucleic Acids Res 19:3055-60). Additionally, one can use PCR, nested primers and PROMOTERFINDER libraries to "walk in" genomic DNA (Clontech). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes. A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region.

The polynucleotides and oligonucleotides of the invention can also be prepared by solid-phase methods, according to known synthetic methods. Typically, fragments of up to about 100 bases are individually synthesized, then joined to form continuous sequences up to several hundred bases.

B4. Applications of Polynucleotides

The polynucleotides and oligonucleotides of the invention have a variety of uses in (1) synthesis of P2 polypeptides, (2) diagnostics, (3) screening for genetic polymorphisms, (4) gene mapping, and (5) therapeutics.

(1) Synthesis of P2 polypeptides In accordance with the present invention, polynucleotide sequences which encode P2, splice isoforms, fragments of the polypeptide, fusion proteins, or functional equivalents thereof, collectively referred to herein as "P2", may be used in recombinant DNA molecules that direct the expression of P2 in appropriate host cells. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode the same, substantially the same, or a functionally equivalent, amino acid sequence may be used to clone and express P2.

As will be understood by those of skill in the art, it may be advantageous to produce P2-encoding nucleic acid sequences possessing non-naturally occurring codons. Codons preferred by a particular prokaryotic or eukaryotic host (Murray, E. *et al.* (1989) Nucl. Acids Res. 17:477-508) can be selected, for example, to increase the rate of P2 polypeptide expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence.

The polynucleotide sequences of the present invention can be engineered in order to alter a P2 coding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression of the gene product. For example, alterations may be introduced using techniques which are well known in the art, *e.g.*, site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, etc.

The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are also described in Sambrook, *et al.*, (*supra*).

The present invention also relates to host cells which are genetically engineered with vectors of the invention, and the production of proteins and polypeptides of the invention by recombinant techniques. Host cells are genetically engineered (i.e., transduced, transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the P2 gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art.

The polynucleotides of the present invention may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host. The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and related sub-cloning procedures are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate transcription control sequence (promoter) to direct mRNA synthesis. Examples of such promoters include: LTR or SV40 promoter, the *E. coli* lac or trp promoter, the phage lambda PL promoter, and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation, and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

The vector containing the appropriate DNA sequence as described above, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein. Examples of appropriate expression hosts include: bacterial cells, such as *E. coli*, *Streptomyces*, and *Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila* and *Spodoptera* Sf9; mammalian cells such as CHO, COS, BHK, HEK 293 or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein. The invention is not limited by the host cells employed.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for P2. For example, when large quantities of P2 are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as Bluescript(R) (Stratagene), in which the P2 coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster (1989) J Biol Chem 264:5503-5509); pET vectors (Novagen, Madison WI); and the like.

In the yeast *Saccharomyces cerevisiae* a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel *et al.* (supra) and Grant *et al.* (1987; Methods in Enzymology 153:516-544).

In cases where plant expression vectors are used, the expression of a sequence encoding P2 polypeptide may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson *et al.* (1984) Nature 310:511-514) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu *et al.* (1987) EMBO J 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi et al (1984) EMBO J 3:1671-1680; Broglie *et al.* (1984) Science 224:838-843); or heat shock promoters (Winter J and Sinibaldi RM (1991) Results. Probl. Cell Differ. 17:85-105) may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. For reviews of such techniques, see Hobbs S or Murry

LE (1992) in McGraw Hill Yearbook of Science and Technology, McGraw Hill, New York, N.Y., pp 191-196; or Weissbach and Weissbach (1988) Methods for Plant Molecular Biology, Academic Press, New York, N.Y., pp 421-463.

P2 may also be expressed in an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* Sf9 cells or in *Trichoplusia* larvae. The P2 coding sequence is cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of P2 coding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect *S. frugiperda cells* or *Trichoplusia* larvae in which P2 is expressed (Smith *et al.* (1983) J Virol 46:584; Engelhard EK *et al.* (1994) Proc Nat Acad Sci 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a P2 coding sequence may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing P2 in infected host cells (Logan and Shenk (1984) Proc Natl Acad Sci 81:3655-3659). In addition, transcription enhancers, such as the rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of a P2 coding sequence. These signals include the ATG initiation codon and adjacent sequences. In cases where P2 coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf D *et al.* (1994) Results Probl Cell Differ 20:125-62; Bittner *et al.* (1987) Methods in Enzymol 153:516-544).

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., and Battey, I. (1986) Basic Methods in Molecular Biology). Cell-free translation systems can also be employed to produce polypeptides using RNAs derived from the DNA constructs of the present invention.

A host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO, HeLa, BHK, MDCK, 293, WI38, etc. have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express P2 may be transformed using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clumps of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M *et al.* (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy I *et al.* (1980) Cell 22:817-23) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler M *et al.*

(1980) Proc Natl Acad Sci 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin F *et al.* (1981) J Mol Biol 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman SC and RC Mulligan (1988) Proc Natl Acad Sci 85:8047-51). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate, GUS, and luciferase and its substrates, luciferin and ATP, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes CA et. al (1995) Methods Mol Biol 55:121-131).

Host cells transformed with a nucleotide sequence encoding P2 may be cultured under conditions suitable for the expression and recovery of the encoded polypeptide from cell culture. The polypeptide produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding P2 can be designed with signal sequences which direct secretion of P2 polypeptide through a prokaryotic or eukaryotic cell membrane.

P2 may also be expressed as a recombinant protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle, WA). The inclusion of a protease-cleavable polypeptide linker sequence between the purification domain and P2 is useful to facilitate purification. One such expression vector provides for expression of a fusion protein compromising P2 (*e.g.*, a soluble P2 fragment) fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography, as described in Porath *et al.* (1992) Protein Expression and Purification 3:263-281) while the enterokinase cleavage site provides a means for isolating P2 from the fusion protein. pGEX vectors (Promega, Madison, WI) may also

be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (e.g., glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, or other methods, which are well know to those skilled in the art.

P2 can be recovered and purified from recombinant cell cultures by any of a number of methods well known in the art, including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

(2) Diagnostic applications. The polynucleotides of the present invention may be used for a variety of diagnostic purposes. The polynucleotides may be used to detect and quantitate expression of P2 or its splice isoforms in patients' cells, for example, in biopsied tissues, bronchial lavage fluid, or blood, by detecting the presence of mRNA encoding P2 or its splice isoforms. A preferred methodology for detecting expression of P2-encoding sequences is the method described in Section I above, particularly using RT-PCR or probe hybridization methods to enhance detection. The diagnostic method can be used to distinguish between absence, presence, and excess expression of P2, and to monitor levels of P2 expression during therapeutic intervention. This method can also be used in a prognostic capacity to assess a subject's risk of developing a disease or disorder associated with expression of P2, including, but not limited to, asthma.

The invention also contemplates the use of the polynucleotides as a diagnostic for diseases resulting from defective P2 genes. These genes can be detected by

comparing the sequences of the defective (i.e., mutant) P2 gene with that of a normal P2 gene. Association of a mutant P2 gene with abnormal P2 biological activity may be verified. In addition, mutant P2 genes can be inserted into a suitable vector for expression in a functional assay system as yet another means to verify or identify mutations. Once mutant genes have been identified, one can then screen populations of interest for carriers of the mutant gene.

Individuals carrying mutations in the gene of the present invention may be detected at the DNA level by a variety of techniques. Nucleic acids used for diagnosis may be obtained from a patient's cells, including but not limited to such as from blood, urine, saliva, placenta, tissue biopsy and autopsy material. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki, *et al.* (1986) Nature 324:163-166) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid of the present invention can be used to identify and analyze mutations in the P2 gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype.

Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA of the invention or alternatively, radiolabeled antisense DNA sequences of the invention. Sequence changes at specific locations may also be revealed by nuclease protection assays, such RNase and S1 protection or the chemical cleavage method (e.g. Cotton, *et al.* (1985) Proc. Natl. Acad. Sci. USA 85:4397-4401), or by differences in melting temperatures. "Molecular beacons" (Kostrikis L.G. *et al.* (1998) Science 279:1228-1229), hairpin-shaped, single-stranded synthetic oligonucleotides containing probe sequences which are complementary to the nucleic acid of the present invention, may also be used to detect point mutations or other sequence changes as well as monitor expression levels of P2. Such diagnostics would be particularly useful for prenatal testing.

Another method for detecting mutations uses two DNA probes which are designed to hybridize to adjacent regions of a target, with abutting bases, where the region of known or suspected mutation(s) is at or near the abutting bases. The two probes may be joined at the abutting bases, e.g., in the presence of a ligase enzyme, but only if both probes are correctly base paired in the region of probe junction. The

presence or absence of mutations is then detectable by the presence or absence of ligated probe.

Also suitable for detecting mutations in the P2 coding sequence are oligonucleotide array methods based on sequencing by hybridization (SBH), as described, for example, in U.S. Patent No. 5,547,839. In a typical method, the DNA target analyte is hybridized with an array of oligonucleotides formed on a microchip. The sequence of the target can then be "read" from the pattern of target binding to the array.

(3) Screening for Genetic Polymorphisms. The polynucleotides of the present invention may be used to screen populations of individuals for heritable genetic variations (polymorphisms) in the P2 gene. The most common type of genetic variations, single nucleotide polymorphisms (SNPs), are single base substitutions (and, less often, insertions or deletions) which occur on an average of 1 variable site every 500-1000 bp of genomic sequence (Nickerson, D.A. (1998) Nat. Genet. 19(3):233-40; Wand, D.G. *et al.* (1998) Science 280:1077-1082). Polymorphisms within the P2 nucleic acid sequences of the present invention may be detected using the methods described in section (2) above, or by other methods known in the art. Genetic polymorphisms within the P2 gene may underlie differential responses to P2- related therapeutics, and are useful as prognostic indicators of a subject's risk of developing a disease or disorder associated with P2, such as asthma.

(4) Gene Mapping. The sequences of the present invention are also valuable for chromosome identification. The P2 sequence has been localized to the 5q31 region of chromosome 5. There is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in the OMIM database (Center for Medical Genetics, Johns Hopkins University, Baltimore, MD and National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD). The OMIM gene map presents the cytogenetic map location of disease genes and other expressed

genes. The OMIM database provides information on diseases associated with the chromosomal location. Such associations include the results of linkage analysis mapped to this interval, and the correlation of translocations and other chromosomal aberrations in this area with the advent of polygenic diseases, such as cancer.

(5) Therapeutic applications. P2 polynucleotides as described above, or their complements, may also be used for therapeutic purposes. Expression of P2 may be modulated through antisense technology, which controls gene expression through complementary polynucleotides, i.e. antisense DNA or RNA, to the control, 5' or regulatory regions of the gene encoding P2. For example, the 5' coding portion of the polynucleotide sequence which codes for the polypeptide of the present invention is used to design an antisense oligonucleotide of from about 15 to 40 base pairs in length. Oligonucleotides derived from the transcription start site, e.g. between positions -10 and +10 from the start site, are preferred. An antisense DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (Lee *et al.* (1979) Nucl. Acids Res. 6:3073; Cooney *et al.* (1988) Science 241:456; and Dervan *et al.* (1991) Science 251: 1360), thereby preventing transcription and the production of P2. An antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of the mRNA molecule into P2 protein (Okano (1991) J. Neurochem. 56:560). The antisense constructs can be delivered to cells by procedures known in the art such that the antisense RNA or DNA may be expressed *in vivo*.

The therapeutic polynucleotides of the invention may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The polypeptides, and agonist and antagonist compounds which are polypeptides, may also be employed in accordance with the present invention by expression of such polypeptides in vivo, which is often referred to as "gene therapy." Cells from a patient may be engineered with a polynucleotide (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being provided to a patient to be treated with the polypeptide. Such methods are well-known in the art. For example,

cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding the polypeptide of the present invention may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells *in vivo* after combination with a suitable delivery vehicle.

Retroviruses from which the retroviral plasmid vectors mentioned above may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus.

The retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the PE501, PA317, psi-2, psi-AM, PA12, T19-14X, VT-19-17-H2, psi-CRE, psi-CRIP, GP+E-86, GP+envAm12, and DAN cell lines as described in Miller (1990; Human Gene Therapy, Vol. 1, pg. 5-14). The vector may transduce the packaging cells through any means known in the art. Such means include, but are not limited to, electroporation, the use of liposomes, and $CaPO_4$ precipitation. In one alternative, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

The producer cell line generates infectious retroviral vector particles which include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed, to transduce eukaryotic cells, either in vitro or in vivo. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

The genes introduced into cells may be placed under the control of inducible promoters, such as the radiation-inducible Egr-1 promoter, (Maceri, H.J., *et al.* (1996) Cancer Res 56(19):4311), to stimulate P2 production or antisense inhibition in response to radiation, *e.g.*, radiation therapy for treating tumors.

C. P2 Polypeptides

From analysis of SEQ ID NO:2, human P2 (also designated herein as isoform P2-A) is a 77-amino acid polypeptide which contains a predicted N-terminal signal peptide from amino acids 1 to about 22. The P2 polypeptide SEQ ID NO:2 contains two regions of basic amino acids: KRRPRR at positions 32-37 and RKRR from positions 66-69, and a cysteine-rich region near the C-terminus containing four cysteines at positions 54, 60, 65, and 70. The mouse homolog of P2, identified herein as SEQ ID NO:7, is 76 amino acids in length.

Fig. 7 shows an amino acid sequence alignment between P2 from human (SEQ ID NO:2) and from mouse (SEQ ID NO:7) prepared using the CLUSTAL-W (ver. 1.4) alignment program in pairwise mode, using the default pairwise alignment parameters (Open gap penalty = 10.0, Extend gap penalty = 0.1, and Blosum30 Similarity Matrix). As shown in the alignment, human and mouse P2 share about 84% amino acid sequence identity (65 identical residues) over the entire polypeptide, and about 95% sequence identity over the 55 amino acid mature polypeptide sequence. Similarity searching of the public sequence databases yielded no polypeptide sequence having significant similarity to the human P2 polypeptide sequence SEQ ID NO:2 or the mouse P2 polypeptide sequence SEQ ID NO:7. Based on its small size, tissue distribution, presence of a signal peptide, cysteine sequence motif, and abundance of basic residues, P2 appears to be a member of a novel chemokine family, denoted CX5C, associated with immune function.

The substantially purified P2 of the invention includes a polypeptide comprising a sequence having at least 80%, preferably at least 85%, 90% or 95% identity to the sequence identified as SEQ ID NO:2, or to a splice isoform of P2, such as the isoform identified as SEQ ID NO:4. Other splice isoform polypeptides of P2 are contemplated, as described above. The polypeptide may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide, preferably a recombinant polypeptide. The polypeptide may be in mature and/or modified form, as defined above. Also

contemplated are fragments having at least 10 contiguous amino acid residues, preferably at least 20 residues, derived from P2 or its isoforms.

The sequence variations are preferably those that are considered conserved substitutions, as defined above. Thus, for example, a polypeptide with a sequence having at least 80% sequence identity with the polypeptide identified as SEQ ID NO:2 (77 amino acid residues) contains up to 15 amino acid substitutions, preferably conserved substitutions as defined above. In a more specific embodiment, the polypeptide has or contains a sequence substantially identical (98-100% identical) to SEQ ID NO:2 or SEQ ID NO:4. P2 may be (i) a polypeptide in which one or more of the amino acid residues in a sequence listed above are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue), or (ii) a polypeptide in which one or more of the amino acid residues includes a substituent group, or (iii) a polypeptide in which P2 is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol (PEG)), or (iv) a polypeptide in which additional amino acids are fused to P2, or (v) an isolated P2 fragment as defined above. Such fragments, variants and derivatives are deemed to be within the scope of those skilled in the art from the teachings herein.

C1. Preparation of P2

Recombinant methods for producing and isolating P2, splice isoforms, and fragments are described above.

In addition to recombinant production, fragments and portions of P2 may be produced by direct peptide synthesis using solid-phase techniques (cf Stewart *et al.* (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield J (1963) J Am Chem Soc 85:2149-2154). *In vitro* peptide synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.) in accordance with the instructions provided by the manufacturer. Portions of P2 may be chemically synthesized separately and combined using chemical methods.

The protein may also be obtained by isolation from natural sources, e.g., by affinity purification using the anti-P2 antibody described in the section below. Furthermore, fragments corresponding to domains, such as the C-terminal cysteine rich region of P2, may be isolated using limited proteolysis techniques known to those of

skill in the art. The amino acid sequence of the fragment(s) so obtained may be used to design nucleotide coding sequence for recombinant production of the fragment(s).

C2. Applications of P2

The P2 polypeptide of the invention has uses in (1) therapeutic treatment methods and compositions and (2) drug screening.

(1) Therapeutic Uses and Compositions The P2 polypeptides of the present invention appear to represent a new class of secreted chemokine. P2 polypeptides may be employed for modulation (*i.e.*, activation or inhibition) of extracellular signaling processes, such as those involved in immune function. In particular, compositions of the present invention may be useful in treating conditions relating to immune system hypersensitivity. Of the four types of immune system hypersensitivity, Type I, or immediate/anaphylactic hypersensitivity, may be particularly amenable to treatment using P2 polypeptides of the present invention. Type I hypersensitivity is due to mast cell degranulation in response to an allergen (*e.g.*, pollen), and includes asthma, allergic rhinitis (hay fever), urticaria (hives), anaphylactic shock, and other illnesses of an allergic nature.

P2 polypeptide compositions are tested in appropriate *in vitro* and *in vivo* animal models of disease, to confirm efficacy, tissue metabolism, and to estimate dosages, according to methods well known in the art.

P2 compositions may be administered by any of a number of routes and methods designed to provide a consistent and predictable concentration of compound at the target organ or tissue. The polypeptide compositions may be administered alone or in combination with other agents, such as stabilizing compounds, and/or in combination with other pharmaceutical agents such as drugs or hormones.

P2 compositions may be administered by a number of routes including, but not limited to oral, intravenous, intramuscular, transdermal, subcutaneous, topical, sublingual, or rectal means. P2 compositions may also be administered via liposomes. Such administration routes and appropriate formulations are generally known to those of skill in the art.

For example, the polypeptide may be given topically to the skin or epithelial linings of body cavities, for infections in such regions. Examples of treatable body

cavities include the vagina, the rectum and the urethra. Conveniently, the polypeptide can be formulated into suppository form for administration to these areas.

The polypeptide can be given via intravenous or intraperitoneal injection. Similarly, the polypeptide may be injected to other localized regions of the body. The polypeptide may also be administered via nasal insufflation. Enteral administration is also possible. For such administration, the polypeptide should be formulated into an appropriate capsule or elixir for oral administration, or into a suppository for rectal administration.

The foregoing exemplary administration modes will likely require that the polypeptides be formulated into an appropriate carrier, including ointments, gels, suppositories. Appropriate formulations are well known to persons skilled in the art.

Dosage of the polypeptide will vary, depending upon the potency and therapeutic index of the particular polypeptide selected. These parameters are easily determinable by the skilled practitioner. As an example, if the polypeptide inhibits mast cell degranulation *in vitro* at a given concentration, the practitioner will know that the final desired therapeutic concentration will be this range, calculated on the basis of expected biodistribution. An appropriate target concentration is in the ng/kg to low mg/kg range, e.g., 50 ng/kg to 1 mg/kg body weight, for IV administration.

A therapeutic composition for use in the treatment method can include the polypeptide in a sterile injectable solution, the polypeptide in an oral delivery vehicle, or the polypeptide in a nebulized form, all prepared according to well known methods. Such compositions comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof.

(2) Screening Methods The present invention includes a method for identifying and isolating coding sequences for proteins which bind to or interact with P2. Such P2-binding proteins may include any proteins which specifically interact with P2, such as a cell-surface P2 receptor. One method contemplated for identification of P2-binding proteins is the "two hybrid" method (see, for example, U.S. Patent Number 5,283,317; Zervos, *et al.* (1993) Cell 72:223-232; Madura *et al.* (1993) J. Biol. Chem. 268:12046-12054; Bartel *et al.* (1993) Biotechniques 14:920-924; Frederickson (1998) Curr. Opin. Biotechnol. 9:90-96). The two hybrid assay method uses the restoration of

transcriptional activation by reconstituting *in vivo* a functional transcription activator complex from two separate fusion proteins. A first hybrid gene, encoding a "bait" fusion protein, comprises a coding sequence for a DNA-binding domain of a transcriptional activator fused in-frame to the coding sequence for a P2 polypeptide. A second hybrid gene, encoding a "sample" fusion protein, comprises a coding sequence for a transcriptional activation domain fused in-frame to a sample coding sequence from a cDNA library. If the bait and sample fusion proteins interact, e.g., forming a P2-dependent complex, they bring into close proximity the DNA binding domain and the activation domain of the transcription activator complex. This proximity is sufficient to cause transcription of a reporter gene which is operatively linked to a transcriptional regulatory region responsive to the transcription activator complex, and expression of the reporter gene can be detected and used to score for the interaction of P2 and sample proteins. Such assays may be performed using commercially available kits, such as the MATCHMAKER Two-Hybrid Systems (Clontech, Cat Nos. K1602-1, K1604-1 and K1605-1).

P2, its binding, catalytic, or immunogenic fragments, or oligopeptides thereof, can be used for screening therapeutic compounds in any of a variety of drug screening techniques. Screening methods which employ P2 for identifying agents which modulate the activity of P2 also form part of the invention. The polypeptide employed in such a test may be membrane-bound, free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between P2 and the agent being tested may be measured, *e.g.* by surface plasmon resonance (BIACORE®; Biacore AB, Uppsala Sweden).

The screening method may include screening for modulating agents which are inhibitors, or alternatively potentiators, of an interaction between P2 and, for example, a target protein which binds P2, such as a P2 receptor. One exemplary method includes the steps of (a) combining P2, or a bioactive fragment thereof, with a target protein and a test modulating agent, under conditions where, but for the test modulating agent, the P2 and the target protein are able to interact; and (b) detecting the formation of a complex which includes the P2 and the target protein either by directly quantitating the complex, or by measuring the bioactivity of the P2 polypeptide or the target protein. A statistically significant change, such as for example a decrease, in the interaction of the P2 and the target protein in the presence of the test modulating agent, relative to what is

detected in the absence of the test modulating agent, is indicative of a modulation (e.g., inhibition or potentiation) of the interaction between P2 and the target protein. The test modulating agent may be a component of a combinatorial library of test agents. The invention also includes, in a related aspect, a modulating agent identified by the screening methods described above. Such modulating agents may include synthetic drugs, antibodies, peptides, or other compounds. Such agents are useful in the treatment of diseases and conditions associated with activation or reduction of P2 activity, including those described under the heading in Section C2 above.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the P2 protein is described in detail by Geysen in PCT Application WO 84/03564, published on Sep. 13, 1984. In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with P2, and washed. Bound P2 is then detected by methods well known in the art. Substantially purified P2 can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

Antibodies to P2, as described in Section IV. below, may also be used in screening assays according to methods well known in the art. For example, a "sandwich" assay may be performed, in which an anti- P2 antibody is affixed to a solid surface such as a microtiter plate and solubilized P2 is added. Such an assay can be used to capture compounds which bind to P2. Alternatively, such an assay may be used to measure the ability of compounds to interfere with the binding of P2 to a P2-binding target protein (e.g., receptor).

D. Anti-P2 Antibodies

In still another aspect of the invention, purified P2 is used to produce anti-P2 antibodies which have diagnostic and therapeutic uses related to the activity, distribution, and expression of P2.

Antibodies to P2 may be generated by methods well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, humanized, single chain, Fab fragments and fragments produced by an Fab expression

library. Antibodies such as those which block ligand binding, are especially preferred for therapeutic use.

P2 for antibody induction does not require biological activity; however, the protein fragment or oligopeptide must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least 4 amino acids, preferably at least 10 or at least 20 amino acids derived from P2. Preferably they should mimic a portion of the amino acid sequence of the natural polypeptide and may contain the entire amino acid sequence of P2. Short stretches of a P2 polypeptide may be fused to a carrier substance such as keyhole limpet hemocyanin, bovine serum albumin, or the like, typically using a bifunctional coupling agent according to well known methods (see Harlow et al., Antibodies: A Laboratory Manual, Cold Spring Harbor Lab (1988, pp. 77-87; and Wong, S.S., Chemistry of Protein Conjugation and Cross-Linking, CRC Press, Boca Raton, Florida (1991)).

For the production of antibodies, various hosts including goats, rabbits, sheep, rats, mice, etc may be immunized by injection with P2 or any portion, fragment or oligopeptide which retains immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are potentially useful human adjuvants.

Polyclonal antibody preparation is by conventional techniques, including injection of the immunogen into a host, such as a rabbit, sheep or mouse, according to immunological protocols generally known in the art (e.g., Harlow, pp. 93-115). Typically, rabbits are injected subcutaneously with the immunogen in an adjuvant, and booster immunizations are given by subcutaneous or intramuscular injection every 2-3 weeks; mice may be injected intraperitoneally according to a similar schedule. Blood is collected at intervals, e.g. 1-2 weeks after each immunization injection. Antisera may be titrated to determine antibody titre with respect to the P2 immunogen, according to standard immuno-precipitation methods (Harlow, pp. 423-470). Successful preparation of polyclonal antibodies against two different peptide sequences from P2 is described in Example 5 below.

Monoclonal antibodies to P2 may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein (1975; Nature 256:495-497), the human B-cell hybridoma technique (Kosbor *et al.* (1983) Immunol Today 4:72; Cote *et al.* (1983) Proc Natl Acad Sci 80:2026-2030) and the EBV-hybridoma technique (Cole, *et al.* (1984) Mol. Cell Biol. 62:109-120). In brief, an immunogen of the type above is used to immunize a host animal, such as a mouse, from which antigen-specific lymphocytes can be obtained for immortalization. An exemplary host is the "autoimmune" MRL/MpJ-lpr mouse available from Jackson Laboratory (Bar Harbor, MN). The hybridoma cell line is grown in a suitable medium (Harlow, pp. 247-270), and monoclonal antibodies ("Mabs") are harvested from the medium and stored according to published methods (Harlow pp. 271-318).

Anti-P2 antibodies may further comprise humanized antibodies or human antibodies. Humanized forms of non-human (e.g., murine) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. Humanized antibodies include human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies may also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-329 (1988); and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source which is non-human. These non-human amino acid residues are often referred to as "import" residues, which are typically taken from an "import" variable domain. Humanization can be essentially performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. Accordingly, such "humanized" antibodies are chimeric antibodies (U.S. Patent No. 4,816,567), wherein substantially less than an intact human variable domain has been substituted by the corresponding sequence from a non-human species. In practice, humanized antibodies are typically human antibodies in which some CDR residues and possibly some FR residues are substituted by residues from analogous sites in rodent antibodies.

Human antibodies can also be produced using various techniques known in the art, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). The techniques of Cole et al. and Boerner et al. are also available for the preparation of human monoclonal antibodies (Cole et al., Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, p. 77 (1985) and Boerner et al., J. Immunol., 147(1):86-95 (1991)).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi *et al.* (1989; Proc Natl Acad Sci 86:3833-3837), and Winter G and Milstein C (1991; Nature 349:293-299).

Antibody fragments which contain specific binding sites for P2 may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse WD *et al.* (1989) Science 256:1275-1281).

D1. Diagnostic Uses of Anti-P2 Antibodies

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the formation of complexes between P2 and its specific antibody and the measurement of complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two noninterfering epitopes on P2 is preferred, but a competitive binding assay may also be employed. These assays are described in Maddox DE *et al.* (1983, J Exp Med 158:1211).

Antibodies which specifically bind P2 are useful for the diagnosis of conditions or diseases characterized by expression of P2. Alternatively, such antibodies may be used in assays to monitor patients being treated with P2, or with compounds which modulate P2 expression. Diagnostic assays for P2 polypeptides include methods utilizing the antibody and a label to detect P2 in extracts of cells, tissues, or biological fluids such as sera. The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, the polypeptides and antibodies will be labeled by joining them, either covalently or noncovalently, with a reporter molecule. A wide variety of reporter molecules are known in the art.

A variety of protocols for measuring P2, using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA; see Example 5), radioimmunoassay (RIA), and fluorescent activated cell sorting (FACS). As noted above, a two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on P2 is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox, *et al.* (*supra*). Such protocols provide a basis for diagnosing altered or abnormal levels of P2 expression. Normal or standard values for P2 expression are established by combining cell extracts taken from normal subjects, preferably human, with antibody to P2 under conditions suitable for complex formation which are well known in the art. The amount of standard complex formation may be quantified by various methods, preferably by photometric methods. Then, standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by

disease. Deviation between standard and subject values establishes the presence of disease state.

The antibody assays are useful to determine the level of P2 present in a particular tissue or body fluid, e.g., biopsied lung tissue, bronchial lavage tissue, blood, urine, or saliva, as an indication of whether P2 is being overexpressed or underexpressed in the tissue, or as an indication of how P2 levels are responding to drug treatment.

D2. Therapeutic Uses of Anti-P2 Antibodies

Therapeutic value may be achieved by administering an antibody specific for P2. Such antibodies may be useful in, for example, modulating inflammation in tissues which overexpress P2. Anti-P2 antibodies may be useful in therapeutic applications including those outlined in section C2 above.

The antibody employed is preferably a humanized monoclonal antibody, or a human Mab produced by known globulin-gene library methods. The antibody is administered typically as a sterile solution by IV injection, although other parenteral routes may be suitable. Typically, the antibody is administered in an amount between about 1-15 mg/kg body weight of the subject. Treatment is continued, e.g., with dosing every 1-7 days, until a therapeutic improvement is seen. In one exemplary method, the antibody treatment is used to decrease the severity of an asthmatic attack, by administering the antibody by inhalation upon presentation of symptoms.

The present invention is further illustrated by way of the following examples, which are not intended to limit the scope of the invention.

Example 1

Isolation of CD4-Positive Cells

A. CD4-positive cells from human peripheral blood. Ficoll-Paque gradient centrifugation was used to enrich for lymphocytes from human peripheral blood samples from asthmatic and non-asthmatic subjects. Briefly, peripheral blood prewarmed to room temperature was diluted 1:3 with phosphate buffered saline without Mg^{+2} and Ca^{+2} (PBS, from GibcoBRL, Life Technologies, NY). 10 mL Ficoll-Paque (Amersham-Pharmacia Biotech) was added with care to the bottom of a 50 mL conical

tube containing 30 mL of the diluted blood to avoid mixing the Ficoll and blood. The sample was spun in a Beckman desktop GPR centrifuge at 2200 rpm for 20 min at room temperature. The white cell interface that formed between the Ficoll and serum layers was collected with a glass pipette. If the collected cell mixture was reddish in color, indicating substantial contamination with red blood cells (RBCs), a RBC lysis step was carried out. In this case, five to ten volumes of ice cold lysis buffer (0.155 M NH_4Cl, 0.01 M $KHCO_3$, 0.01 mM EDTA) was mixed with cell interface and incubated for 15 min on ice. After centrifugation at 1800 rpm for 10 min, the supernatant was discarded.

The cells were washed twice by addition of 40 mL PBS and spun at 1800 rpm for 10 min. The supernatant was discarded. The cell pellet, mainly lymphocytes, was resuspended with a concentration of 1.25×10^8 cells per mL in PBS with 2 mM EDTA and 0.5% BSA (PBS-buffer) precooled at 4°C.

The lymphocyte preparation was mixed with magnetic beads coupled with anti-CD4 antibody in a ratio of 20 μL beads to 80 μl cell suspension containing $1x10^7$ cells as described in the protocol (Miltenyi Biotech, Inc., Auburn, CA). The cell-bead mixture was incubated at 6-12°C for 15 min. The cells were then washed by adding a 10x volume of PBS-buffer followed by centrifugation at 1800 rpm for 10 min at 4°C. The pelleted cells were fully resuspended in 0.5 mL of PBS buffer before passing through a prefilter (Miltenyi Biotech, Inc.) to filter debris. CD4+ cells were isolated using magnetic column sorting (RS+ column with pore size of 30 nm) according to the protocol supplied (Miltenyi Biotech, Inc.).

B. CD4-positive cells from broncho-alveolar lavage (BAL). BAL samples from asthmatic and nonasthmatic human patients were spun down at 1800 rpm for 10 min. The cells were resuspended in 50 mL Hank's balanced salt solution in two T150 tissue culture flasks and incubated at 37°C for 2 hours. The majority of the cells adhered to the surface of the flasks and were discarded (macrophages). The cells in the medium were separated from debris by filtering through two layers of cheese cloth. The cells were centrifuged and washed twice with PBS before they were labeled with anti-CD4 antibody coated magnetic beads. The isolation procedure was essentially the same as for PBCs above except that a column with larger pore size (100 nm) was used in the magnetic column sorting step.

Example 2

Isolation of Total Cell RNA from CD4-Positive Cells

Total cell RNA was isolated from CD4-positive cells (Example 1) using a "RNEASY MINI KIT" from Qiagen. In a typical application, the number of cells applied to each filter was between about 5×10^4 and 1×10^7. To 30 μL RNA eluted from a "RNEASY MINI COLUMN" spin column were added 7.5 μL of 5x transcription buffer and 2.0 μL (2.0 U) of RNase-free RQ1 DNAse (Promega). The digestion was allowed to stand at room temperature for 30 min. The DNAse was separated from RNA using a second round of "RNEASY MINI COLUMN" purification. The RNA was eluted in 30 μL water which had been treated with diethylpyrocarbonate (DEPC) and collected in an autoclaved 0.5 mL microcentrifuge tube. About 3 μL of the RNA was electrophoresed on a 1% agarose formaldehyde gel (Amicon Northern Max Kit) to examine the integrity of the RNA isolated and to evaluate its concentration along with UV spectrophotometer measurement (1 OD_{260} = 40 μg/mL).

Example 3

Preparation and Amplification of Double-Stranded DNA

Selection of primer sets. Primers for multiplex amplification of target sequences were selected with the help of computer software known as "PRIMER PREMIER", Version 4 for "WINDOWS", which is commercially available from Premier Biosoft International (Palo Alto, CA, 94303). In brief, full length cDNA sequences for each target polynucleotide were entered as input sequences for primer design. The program was instructed to identify all NlaIII sites in the targets, and to tabulate forward and reverse primer candidates based on the following criteria: (1) the primers must flank each NlaIII site, (2) the primers must have melting temperatures (Tm), with respect to their complementary sequences, within a range of 55°C to 65°C based on the default buffer conditions of the program; (3) the primers must bind to a sequence window within 60 nucleotides upstream or downstream of each NlaIII site. The program produced a list of primer sets for each target, in order of preference determined by the program, from which final primer pairs were selected manually based on the following constraints: (4) the primer pair produces a PCR product 60 to 70 nucleotides in length; (5) the NlaIII site is located at or near the middle of the predicted PCR product; (6) the primers do not hybridize significantly with other primers or target

sequences; (7) the primers do not contain long GC or AT-rich regions, and do not undergo significant internal secondary structure formation; (8) if the target sequence contained more than one NlaIII site, the site closest to the 3'-end of the target mRNA was usually selected, unless the primer set did not adequately satisfy the other criteria noted above.

Primers for 17 representative targets of interest, which were amplified as noted below, are provided in Appendix I.

First strand cDNA synthesis. Total RNA (1-5 μg) from CD4 positive cell was converted to double-stranded (ds) cDNA using SUPERSCRIPT II RT (reverse transcriptase, Life Technologies, Inc) and a mixture of gene-specific-primers, using at least a 10^3 to 10^5-fold excess of each primer over template.

In a specific example, 2 μL of gene-specific primer pairs (116 pairs of gene-specific primers, including for P2 polynucleotide, 4 pmol each) were mixed with 16 μL (15 - 500 ng) RNA. The sample was heated at 70°C for 10 min and transferred to a 55°C water bath. 12 μL of cDNA reaction mix containing 3 μL 10x buffer (200 mM Tris-HCl, pH 8.4, 500 mM KCl), 3 μL 25 mM $MgCl_2$, 3 μL 10 mM dNTPs, and 3 μL 0.1 M DTT was added. The mixture was incubated at 55°C for another 5 min, after which 1.5 μL "SUPERSCRIPT" reverse transcriptase (200 U/μL stock solution, Gibco/BRL, MD) was added, and the transcription reaction proceeded for another 30 min. The reaction was terminated by incubation at 70°C for 10 min.

PCR Amplification. A PCR reaction was set up to amplify the cDNA from the preceding step. The reaction used from half to all of the cDNA material from the preceding step, 10 μL 10x PCR buffer, 10 μL 2 mM dNTPs, 2 μL gene-specific primer pairs (2 pmol each), 70 μL water, and 2 μL Taq polymerase (from Clontech "cDNA ADVANTAGE MIX"). The reaction was heated at 94°C for 2 min to inactivate the anti-Taq antibody followed by 20-25 cycles with 94°C for 30 seconds, 55°C for 30 seconds, and 68°C for 30 seconds. An aliquot of PCR products was examined on a 4% agarose gel while the rest of it was extracted with phenol and precipitated with a DNA recovery agent ("QUICK PRECIP", Edge Biosystems, MD) which gave yields of 90-95%.

Example 4

Measurement of Target Levels

Adaptors. Two double-stranded adaptors having the sequences shown in Fig. 3 and identified herein as SEQ ID NO:8 and SEQ ID NO:9 (Adaptor #1, forward and reverse strands, respectively), and SEQ ID NO:10 and SEQ ID NO:11 (Adaptor #2, forward and reverse strands, respectively) were synthesized by standard automated synthetic methods, where, as shown in Fig. 3, single-underlined sequences (thin underlining) on the left side of each adaptor indicate a first set of forward and reverse primer sequences for the final PCR step used to amplify di-TID multimers (see below), thicker single-underlined sequences (including the double-underlined bases) on the right side of each adaptor indicate a second set of forward and reverse primers that can be used for PCR, double-underlined bases indicate a recognition site for BsmFI, and "p" indicates the presence of a terminal phosphate group.

Restriction Endonuclease Digestion. The amplified cDNA product from Example 3 was resuspended and digested with 50 units of restriction endonuclease NlaIII in a total volume of 50 μL at 37°C for 2 hr. After phenol extraction and precipitation, the digested DNA was dissolved in water and divided into two equal aliquots.

Adaptor Ligation and Optional Truncation and Gel Purification. The two aliquots were subjected to ligation with 25 pmol of either Adaptor #1 or Adaptor #2 (Fig. 3), respectively, in a 20 μL volume for 1 hr using a "RAPID LIGATION PROTOCOL" from Boehringer Mannheim. The sample volumes were then adjusted to 50 μL containing 10 units of BmsFI and corresponding restriction enzyme buffer. The BsmFI digestions proceeded at 65°C overnight with mineral oil overlays on top of samples. After digestion, the two separate aliquots were combined, electrophoresed on a neutral 12% acrylamide gel (Model V16 from BRL, 1.5 mm comb and spacer) for 2.5 hr at 190 V. The gel was stained with SYBR Green I (Molecular Probes, Eugene, OR) for 40 min. The band of about 52-56 bp was excised, cut into small pieces and eluted in 2 mL water at 37°C overnight with shaking. The eluted DNA was dried under vacuum with low heat and precipitated with "QUICK-PRECIP" reagent (supra).

Fill-in Reaction. The resultant fragment mixture was resuspended in 10 μL ligation buffer from a "RAPID LIGATION KIT" containing 1 μL of 2 mM dNTPs, 1 μL (2 units) Klenow polymerase (Boehringer Mannheim). A Klenow fill-in reaction was

performed at room temperature for 30 min after which 10 μL of ligation buffer containing 1 μL of T4 DNA ligase was added. The ligation reaction proceeded at room temperature for 2 hr.

PCR Amplification of TID dimers. The ligated products (TID dimers) were PCR-amplified in a 100 μL reaction volume containing 10 μL of ligation mix from the preceding step, 10 μL PCR buffer (AMPLITAQ, Perkin-Elmer), 10 μL 2 mM dNTPs, 7.5 μL 25 mM $MgCl_2$, 5 μL containing 250 pmol of each forward primer (5'-AGGCTTAATAGGGACAT-3'; SEQ ID NO:12) and reverse primer (5'-ACGATGTACGGGGACAT-3'; SEQ ID NO:13). 20-25 cycles were performed (94°C for 30 seconds, 40°C for 30 seconds, 72°C for 30 seconds). The PCR product was phenol-extracted and recovered by "QUICK PRECIP". The DNA was digested with Nla III in 50 μL containing 50 units of restriction endonuclease for 2 hr at 37°C. The digested DNA fragments were separated by a neutral 14% acrylamide gel, the 28 bp DNA fragment band was excised from the gel and eluted as described previously. The eluted TID dimer mixture was recovered by "QUICK PRECIP" as above, and dried.

TID multimer synthesis. TID dimer units were ligated together to form multimers of approximately 20-25 TID dimer units as follows. A "RAPID DNA LIGATION KIT" (Boehringer Mannheim Cat. No. 1635 379), including supplied buffers, was used in all ligations. If necessary, the 28 bp fragment mixture was concentrated to 4-5 μL in a SpeedVac without heating, before ligation. Typically, not more than about 0.5 μg of TID dimer was used (1/5 of the total 28 bp TID dimer product, based on band intensity in a 4% agarose stained with ethidium bromide).

The ligation reactions, 11 μL each, contained:

2 μL 28 bp TID dimer product (0.5 μg, or about 1/5 of total 28 bp product)
1 μL 10 mM spermidine
1 μL 0.1 pmol/μL adaptor A3
1 μL 5X dilution buffer
5 μL 2X ligation buffer
1 μL DNA ligase

The structure of the double-stranded terminator adaptor (A3, see SEQ ID NO:15 and 16) was as follows:

5'-TTTCTACTGAGTATCTCTAGGATCCTCTCATG-OH-3'
3'-GATGACTCATAGAGATCCTAGGAGA-P-5'

The first three component solutions were carefully mixed together, followed by addition and careful mixing of the 2X ligation buffer (5 μL), and then addition and careful mixing of the DNA ligase (1 μL). The mixture was allowed to react at room temperature for 15 to 60 minutes, and was then incubated at 16°C over night. After this first ligation, an additional capping step was performed by adding 0.5 μL of excess adaptor A3 (10 pmol/μL stock solution) and 0.5 μL of ligase. The ligation mixture was incubated at room temperature for 15 to 60 minutes and then at 16°C over night. Following the second ligation step, a 100-500 bp fraction of the ligated TID dimer material was isolated using agarose gel electrophoresis in 2% agarose gel, and then purified using a QIAEX II or QIAQUICK Gel Extraction kit (Qiagen). As a control, a 100-500 bp fraction of a lambda DNA HindIII/NlaIII digest was included in the gel. A 400-800 bp fraction was also collected using the same techniques.

PCR Amplification of TID multimers. The purified DNA was concentrated to 4 to 10 μL by SpeedVac on low heat, and 2-5 μL of the concentrated DNA was PCR-amplified in a 30 μL solution containing:

5.0 μL DNA fraction 400-800bp.
3.0 μL 10X KLENTAQ amplification buffer
0.33 μL 25 mM dNTP
0.6 μL KlenTaq polymerase
1.5 μL 20 pmol/μL primer PDS1-CUA
Water to final volume of 30 μL

Primer PDS1-CUA has the sequence (SEQ ID NO:14):

5'-CUACUACUACUACTGAGTATCTCTAGGATCCTCTC-3'

wherein U represents uridine residues (ribose forms). Control amplifications were also run on the ligated phage lambda fragments. The PCR amplifications were performed with 22-25 cycles of 94°C for 30 sec, 55°C for 30 sec, and 71°C for 1 min.

Multimer library. A fraction of the amplified DNA having a size range of 400-700 bp was concentrated to 2 - 4 μL, and half of the concentrated material was annealed to pAMP10 vector using the procedure described by the manufacturer ("CLONE AMP" pAMP10 System kit, Life Technologies) in a final volume of 10 μL containing:

1 μL DNA (400-700 bp multimers)
1 μL DNA vector pAMP10
7.5 μL Annealing buffer
0.5 μL UDG

After incubation at 37°C for 30 min, 2 μL of the reaction mixture was used to transform XL1-Blue MR competent cells according to a standard protocol (Stratagene, San Diego). Cells were plated on Amp-containing plates, incubated overnight, and the colonies were picked and grown overnight in 96 well plates with LB and Amp (100 μg/mL) at 37°C. The next day, master plates were replicated and frozen at –70°C after glycerol was added to 15%.

Purification of sequencing template. Plasmid DNA was prepared from plated colonies using QIAPREP 96 TURBO Miniprep kits (Qiagen) according to the manufacture's protocol with minor modifications, followed by sequence analysis using a 377 ABI-Perkin-Elmer sequencer with "TaqFS" polymerase (Perkin-Elmer). Usually, the median clone size was about 380-400 bp, equivalent to approximately 25 to 30 TIDs. About 2,000 to 8,000 TIDs were sequenced per sample. For finer profiling, genes were categorized into four group according to their expression level (high, medium, low, and very low).

Results. CD4-positive cells from BAL and blood samples from three asthmatic and four non-asthmatic patients were isolated, and the expression levels of 116 genes were determined as described above. Fig. 5 (BAL samples) and Fig. 6 (blood samples) show expression levels of 17 of the measured genes. The expression levels of 15 genes appear to be relatively constant between asthmatics and non-asthmatics. For example, it can be seen from Figs. 5 and 6 that the expression of CAML does not vary significantly

between asthmatics and non-asthmatics. However, the expression of IL10, which is known to be associated with asthma, varies considerably between asthmatic and non-asthmatic patients. Fig. 5 shows that IL10 is highly expressed in BAL cells of one asthmatic patient, and Fig. 6 shows high expression of IL10 in blood samples from three of four asthmatic patients. In patient ER, P2 is very highly expressed in both BAL and blood cells. Overall, the expression levels of P2 mirror the expression levels of IL10.

Example 5

Anti-P2 Antibodies

New Zealand white rabbits were immunized with one of the following peptides (2 rabbits per peptide) which had been conjugated to KLH by standard techniques: PEVSSAKRRPRRMTPFWR (pep-1) and CRDDSECITRLCRKRRCS (pep-2), which correspond to residues 26-43 and 54-71 from SEQ ID NO:2 above. The following injection schedule was used: day 0: 200 μg; days 14, 28, 42 and 56: 100 μg. Preimmune serum was collected on day 0, and three subsequent bleeds (25 mL each) were taken on days 49, 63, and 77. Antisera were collected on day 77.

An ELISA was performed as follows, using an SOP-HRP detection format. To separate wells of a 96-well "IMMUNOLON 2" flat bottom polystyrene plate was added 100 μL (200 ng in 0.05 M sodium carbonate, pH 9.5) of peptide-KLH conjugate, followed by incubation for 1 hour at 37°C. The wells were then washed 5 times with PBS (phosphate buffered saline) containing 0.05% "TWEEN 20", and incubated with 200 μL antibody (various dilutions) at 37°C for 1 hour. After washing 5 times with PBS/0.05% "TWEEN 20", the wells were incubated with 100 μL of HRP conjugated goat anti-human IgG at 1:1000 final dilution in antibody diluent for 30 min at 37°C. The wells were washed 5 times with PBS/0.05% "TWEEN 20, and then incubated with 100 μL substrate solution (prepared by mixing 2 OPD-HCl tablets, 5 mg each, 10 μL 30% hydrogen peroxide, and 10 mL 0.1 M sodium citrate, pH 5.0) for 15 minutes at room temperature. The reaction was stopped by adding 2 M sulfuric acid, and the absorbance was read at 490 nm. Antisera from all four rabbits were found to be specific for the two immunogens. For pep-1, the titre was higher than 1:200,000. For pep-2, the titre was at least 1:25,000.

Although the invention has been described with reference to specific embodiments and illustrations, it will be recognized that various modifications can be made without departing from the spirit of the invention as disclosed herein.

Appendix I

Exemplary Targets and Primer Pairs

Name	Gene Description	GenBank Accession Number	primer SEQ ID (forward) SEQ ID (reverse)
A116A	unknown	N/A	SEQ ID NO:17 SEQ ID NO:18
BIGH3	Human transforming growth factor-beta induced gene product	M77349	SEQ ID NO:19 SEQ ID NO:20
CAML	H. sapiens transmembrane activator and CAML interactor	AF023614	SEQ ID NO:21 SEQ ID NO:22
CD4	Human T4 surface glycoprotein CD4	M35160	SEQ ID NO:23 SEQ ID NO:24
CXCR4	H. sapiens chemokine receptor-4	AF025375	SEQ ID NO:25 SEQ ID NO:26
E118	unknown	N/A	SEQ ID NO:27 SEQ ID NO:28
P2	described herein	N/A	SEQ ID NO:29 SEQ ID NO:30
HSP70	Human heat shock protein 70 (hsp70)	L12723	SEQ ID NO:31 SEQ ID NO:32
IL10	Human interleukin 10	U16720	SEQ ID NO:33 SEQ ID NO:34
IRF1	Human interferon regulatory factor 1	L05072	SEQ ID NO:35 SEQ ID NO:36
LynBP	Human Lyn B protein	M79321	SEQ ID NO:37 SEQ ID NO:38
MHCB:GP	Human MHC protein homologous to chicken B complex	M24194	SEQ ID NO:39 SEQ ID NO:40
MIF	Human migration inhibitory factor	M25639	SEQ ID NO:41 SEQ ID NO:42
PKC-B	Human protein kinase C, type beta I	X06318	SEQ ID NO:43 SEQ ID NO:44
STAT5	Human signal transducer and activator of transcription (Stat5A)	U43185	SEQ ID NO:45 SEQ ID NO:46
TAFII55	Human TFIID subunit TAFII55	U18062	SEQ ID NO:47 SEQ ID NO:48
TCF1	H.sapiens T cell factor 1	Z47361	SEQ ID NO:49 SEQ ID NO:50

Claims:

1. A method for simultaneously determining the levels of a plurality of selected target polynucleotide sequences in a sample, comprising:

reacting the sample with a plurality of sequence-selective primer pairs, wherein each primer pair is designed to flank a defined region in a different target sequence of interest, and each defined region contains a selected restriction endonuclease site, to form double-stranded copies of said defined regions in direct proportion to the target levels originally present in the sample,

cleaving the double-stranded copies with one or more endonucleases, to form a cleavage mixture of DNA fragments,

reacting separate first and second aliquots of the cleavage mixture with first and second adaptors, respectively, to form mixtures of first and second adaptor-fragment conjugates, wherein the first and second adaptors each contain (i) a first terminal end that is cohesive with the cleaved termini in the DNA fragments, (ii) a second terminal end that is non-cohesive with respect to the DNA fragments, so that only one end of each adaptor binds to a DNA fragment, and (iii) a unique-sequence segment that allows the first and second adaptors to be distinguished from each other, wherein the resultant conjugates comprise an adaptor segment and a target-identifier segment, by ligation of cohesive ends,

combining said first and second conjugate mixtures under conditions effective to form conjugate dimers containing a conjugate from each of the first and second conjugate mixtures, such that the conjugates in each dimer are joined via their target-identifier segments,

cleaving the adaptors from the conjugate dimers, to produce a plurality of target-identifier dimers,

polymerizing said target-identifier dimers to form dimer multimers, and

quantifying the relative abundances of target-identifiers in one or more dimer multimers to provide an estimate of the levels of the selected target sequences in the sample.

Appendix I

Exemplary Targets and Primer Pairs

Name	Gene Description	GenBank Accession Number	primer SEQ ID (forward) SEQ ID (reverse)
A116A	unknown	N/A	SEQ ID NO:17 SEQ ID NO:18
BIGH3	Human transforming growth factor-beta induced gene product	M77349	SEQ ID NO:19 SEQ ID NO:20
CAML	H. sapiens transmembrane activator and CAML interactor	AF023614	SEQ ID NO:21 SEQ ID NO:22
CD4	Human T4 surface glycoprotein CD4	M35160	SEQ ID NO:23 SEQ ID NO:24
CXCR4	H. sapiens chemokine receptor-4	AF025375	SEQ ID NO:25 SEQ ID NO:26
E118	unknown	N/A	SEQ ID NO:27 SEQ ID NO:28
P2	described herein	N/A	SEQ ID NO:29 SEQ ID NO:30
HSP70	Human heat shock protein 70 (hsp70)	L12723	SEQ ID NO:31 SEQ ID NO:32
IL10	Human interleukin 10	U16720	SEQ ID NO:33 SEQ ID NO:34
IRF1	Human interferon regulatory factor 1	L05072	SEQ ID NO:35 SEQ ID NO:36
LynBP	Human Lyn B protein	M79321	SEQ ID NO:37 SEQ ID NO:38
MHCB:GP	Human MHC protein homologous to chicken B complex	M24194	SEQ ID NO:39 SEQ ID NO:40
MIF	Human migration inhibitory factor	M25639	SEQ ID NO:41 SEQ ID NO:42
PKC-B	Human protein kinase C, type beta I	X06318	SEQ ID NO:43 SEQ ID NO:44
STAT5	Human signal transducer and activator of transcription (Stat5A)	U43185	SEQ ID NO:45 SEQ ID NO:46
TAFII55	Human TFIID subunit TAFII55	U18062	SEQ ID NO:47 SEQ ID NO:48
TCF1	H.sapiens T cell factor 1	Z47361	SEQ ID NO:49 SEQ ID NO:50

Claims:

1. A method for simultaneously determining the levels of a plurality of selected target polynucleotide sequences in a sample, comprising:

reacting the sample with a plurality of sequence-selective primer pairs, wherein each primer pair is designed to flank a defined region in a different target sequence of interest, and each defined region contains a selected restriction endonuclease site, to form double-stranded copies of said defined regions in direct proportion to the target levels originally present in the sample,

cleaving the double-stranded copies with one or more endonucleases, to form a cleavage mixture of DNA fragments,

reacting separate first and second aliquots of the cleavage mixture with first and second adaptors, respectively, to form mixtures of first and second adaptor-fragment conjugates, wherein the first and second adaptors each contain (i) a first terminal end that is cohesive with the cleaved termini in the DNA fragments, (ii) a second terminal end that is non-cohesive with respect to the DNA fragments, so that only one end of each adaptor binds to a DNA fragment, and (iii) a unique-sequence segment that allows the first and second adaptors to be distinguished from each other, wherein the resultant conjugates comprise an adaptor segment and a target-identifier segment, by ligation of cohesive ends,

combining said first and second conjugate mixtures under conditions effective to form conjugate dimers containing a conjugate from each of the first and second conjugate mixtures, such that the conjugates in each dimer are joined via their target-identifier segments,

cleaving the adaptors from the conjugate dimers, to produce a plurality of target-identifier dimers,

polymerizing said target-identifier dimers to form dimer multimers, and

quantifying the relative abundances of target-identifiers in one or more dimer multimers to provide an estimate of the levels of the selected target sequences in the sample.

2. The method of claim 1, wherein said target-identifiers have lengths of from 10 to 50 basepairs.

3. The method of claim 1, wherein said target-identifiers have lengths of from 10 to 20 basepairs.

4. The method of claim 1, wherein said sequenced multimers contain from 20 to 60 target-identifier dimers.

5. The method of claim 1, wherein said sequenced multimers contain at least 400 nucleotides.

6. The method of claim 1, wherein prior to forming said conjugate dimers, the conjugates in said first and second conjugate mixtures are cleaved with a Type IIs restriction endonuclease to shorten the target-identifier segments in the conjugates prior to dimerization.

7. The method of claim 1, wherein said target polynucleotide sequences comprise RNA sequences.

8. The method of claim 7, wherein the sample is a total RNA sample.

9. The method of claim 7, wherein the target sequences are mRNA sequences.

10. The method of claim 9, wherein the mRNA fragments are obtained by solid phase capture on a poly-T oligonucleotide matrix.

11. The method of claim 1, wherein said first restriction endonuclease is NlaIII.

12. The method of claim 1, wherein prior to said combining, a third aliquot of the cleavage mixture is reacted with a third double-stranded adaptor having (i) a first terminal end that is cohesive with the cleaved termini in the DNA fragments, (ii) a second terminal end that is non-cohesive with respect to the DNA fragments, so that

only one end of each adaptor binds to a DNA fragment, and (iii) a unique-sequence segment that allows the third adaptor to be distinguished from the first and second adaptors.

13. A polynucleotide mixture comprising a plurality of dimer multimers prepared as recited in claim 1.

14. The polynucleotide mixture of claim 13, wherein the target-identifiers have lengths of from 10 to 50 basepairs.

15. The polynucleotide mixture of claim 13, wherein the target-identifiers have lengths of from 10 to 20 basepairs.

16. The polynucleotide mixture of claim 13, wherein the multimers contain from 20 to 60 target-identifier segments.

17. The polynucleotide mixture of claim 13, wherein the multimers contain from 400 to 800 base pairs.

18. The polynucleotide mixture of claim 13, wherein the target-identifier sequences are derived from mRNA gene sequences.

19. A method for simultaneously determining expression levels of a plurality of selected target gene sequences in a sample, comprising:

forming a cDNA fragment mixture by treating the sample with a plurality of gene-specific primer pairs, wherein each pair comprises first and second primers which are complementary to opposite ends of a different selected region of a selected target gene, and each selected region encodes a selected restriction endonuclease site,

cleaving the double-stranded copies with one or more endonucleases, to form a cleavage mixture of cDNA fragments,

reacting separate first and second aliquots of the cleavage mixture with first and second adaptors, respectively, to form mixtures of first and second adaptor-fragment conjugates, wherein the first and second adaptors each contain (i) a first terminal end

that is cohesive with the cleaved termini in the DNA fragments, (ii) a second terminal end that is non-cohesive with respect to the DNA fragments, so that only one end of each adaptor binds to a DNA fragment, and (iii) a unique-sequence segment that allows the first and second adaptors to be distinguished from each other, wherein the resultant conjugates comprise an adaptor segment and a target-identifier segment, by ligation of cohesive ends,

combining said first and second conjugate mixtures under conditions effective to form conjugate dimers containing a conjugate from each of the first and second conjugate mixtures, such that the conjugates in each dimer are joined via their target-identifier segments,

cleaving the adaptors from the conjugate dimers, to produce a plurality of target-identifier dimers,

polymerizing said target-identifier dimers to form dimer multimers, and

quantifying the relative abundances of target-identifiers in one or more dimer multimers to provide an estimate of expression levels of the selected target sequences.

20. The method of claim 19, wherein said target-identifiers have lengths of from 10 to 50 basepairs.

21. The method of claim 19, wherein said target-identifiers have lengths of from 10 to 20 basepairs.

22. The method of claim 19, wherein said sequenced multimers contain from 20 to 60 target-identifier segments.

23. The method of claim 19, wherein said sequenced multimers contain from 400 to 800 base pairs.

24. The method of claim 19, wherein prior to forming said conjugate dimers, the conjugates in said first and second conjugate mixtures are cleaved with a Type IIs restriction endonuclease to shorten the target-identifier segments in the conjugates prior to dimerization.

25. The method of claim 19, wherein the sample is a total RNA sample.

26. The method of claim 19, wherein said first restriction endonuclease is NlaIII.

27. The method of claim 19, wherein prior to said combining, a third aliquot of the cleavage mixture is reacted with a third double-stranded adaptor having (i) a first terminal end that is cohesive with the cleaved termini in the DNA fragments, (ii) a second terminal end that is non-cohesive with respect to the DNA fragments, so that only one end of each adaptor binds to a DNA fragment, and (iii) a unique-sequence segment that allows the third adaptor to be distinguished from the first and second adaptors.

28. An isolated and purified P2 polypeptide having an amino acid sequence at least 80% identical to SEQ ID NO:2 or SEQ ID NO:4.

29. The isolated polypeptide of claim 28, having an amino acid sequence at least 95% identical to SEQ ID NO:2 or SEQ ID NO:4.

30. The isolated polypeptide of claim 28, having the amino acid sequence SEQ ID NO:2 or SEQ ID NO:4.

31. An isolated polynucleotide selected from the group consisting of:

(a) a polynucleotide having a sequence which encodes the P2 polypeptide of claim 1, and

(b) a polynucleotide having a sequence complementary to the sequence of (a).

32. The polynucleotide of claim 31, comprising a sequence having at least 80% identity to SEQ ID NO:1 or SEQ ID NO:3.

33. The polynucleotide of claim 31, comprising residues 192-422 of SEQ ID NO:1 or residues 192-311 of SEQ ID NO:3.

34. The polynucleotide of claim 31, having the sequence SEQ ID NO:1 or SEQ ID NO:3.

35. An isolated nucleic acid molecule which hybridizes under high-stringency conditions to the polynucleotide of claim 31, said nucleic acid molecule having a length of at least 15 nucleotides.

36. The isolated nucleic acid molecule of claim 35 having a length of at least 20 nucleotides.

37. An isolated polynucleotide selected from the group consisting of

(a) nucleotides 1047 to 1353 of SEQ ID NO:5,

(b) nucleotides 1354 to 1551 of SEQ ID NO:5,

(c) nucleotides 1552 to 1691 of SEQ ID NO:5,

(d) nucleotides 1692 to 2011 of SEQ ID NO:5,

(e) nucleotides 2012 to 2633 of SEQ ID NO:5, and

(f) the complement of any of (a)-(e).

38. An isolated nucleic acid molecule which hybridizes under high-stringency conditions to the polynucleotide of claim 37, said nucleic acid molecule having a length of at least 15 nucleotides.

39. The isolated nucleic acid molecule of claim 38, having a length of at least 20 nucleotides.

40. An expression vector containing the polynucleotide of claim 33 and regulatory elements effective for expression of the polynucleotide in a suitable host.

41. A purified antibody which specifically binds to the polypeptide of claim 28.

42. A method for detecting a polynucleotide which encodes P2 in a biological sample, comprising the steps of:

(a) hybridizing the molecule of claim 38, under high-stringency conditions, to nucleic acid material of said biological sample, thereby forming a hybridization complex; and

(b) detecting said hybridization complex;

wherein the presence of said hybridization complex correlates with the presence of a polynucleotide encoding P2 in said biological sample.

43. The method of claim 42, wherein the molecule has a length of at least 20 nucleotides.

44. A method of identifying a compound which binds to and modulates the activity of P2, the method comprising:

(a) contacting a test compound with P2,

(b) measuring the effect of the test compound on the activity of P2, and

(c) selecting the test compound as a candidate compound if its effect on the activity of P2 is above a selected threshold level.

45. The method of claim 44, wherein said test compound is a component of a combinatorial library.

46. The method of claim 44, wherein said test compound is a polypeptide.

47. A candidate compound identified by the method of claim 44.

48. A method of identifying a compound which inhibits binding of a P2-binding protein to P2, the method comprising:

(a) incubating together a test compound, P2, and the P2-binding protein, under conditions which allow the P2-binding protein bind to P2 in the absence of said test compound,

(b) measuring the extent of binding of the P2-binding protein to P2 in the presence of the test compound, and

(c) identifying the test compound as a candidate compound if the extent of binding of the P2-binding protein to P2 in the presence of the test compound is less than

the extent of binding of the P2-binding protein to P2 in the absence of the test compound.

49. A candidate compound identified by the method of claim 48.

50. A method for detecting P2 in a biological sample, comprising the steps of:

(a) contacting with said biological sample the antibody of claim 41, thereby forming an antibody-antigen complex; and

(b) detecting said antibody-antigen complex;

wherein the presence of said antibody-antigen complex correlates with the presence of P2 protein in said biological sample.

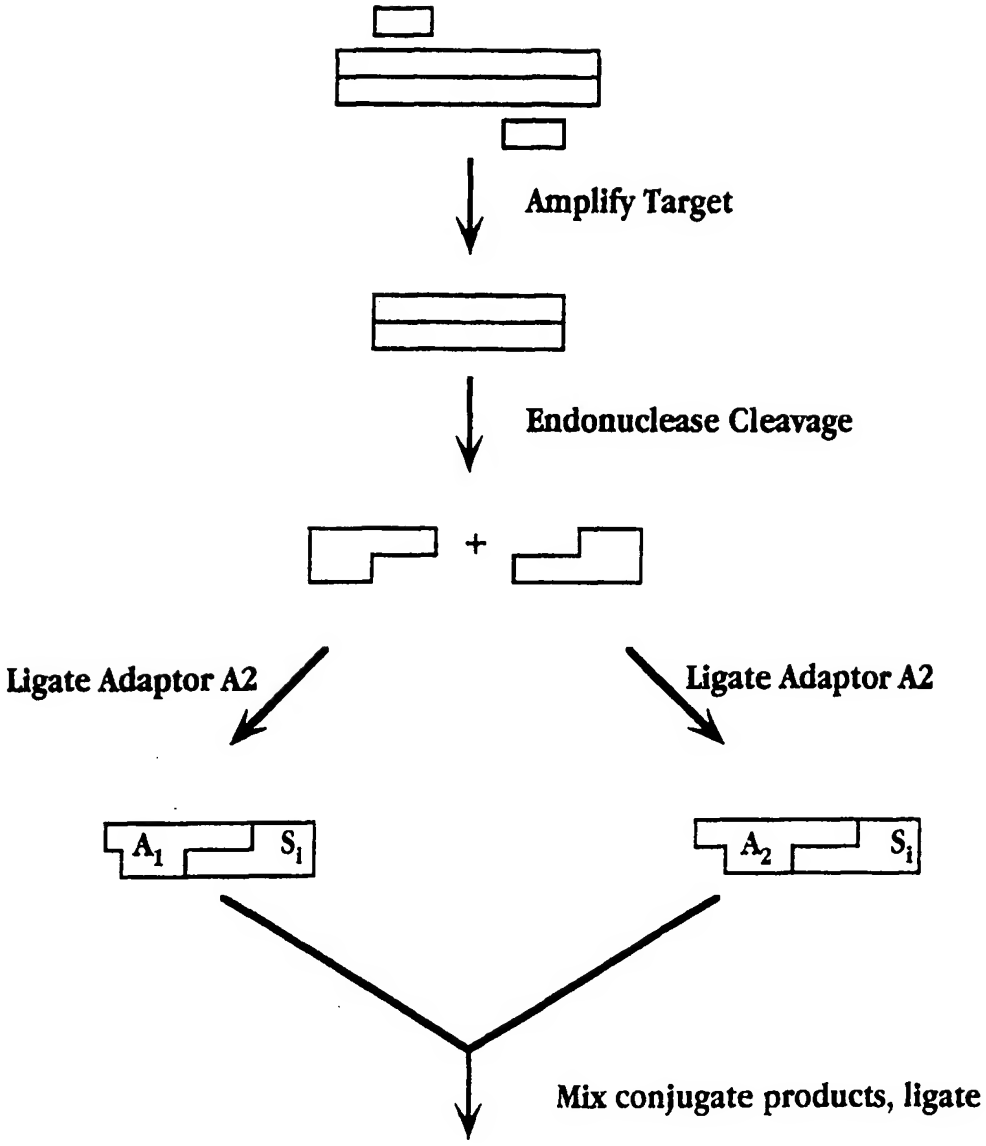

Amplify Target
Endonuclease Cleavage
+
Ligate Adaptor A2
Ligate Adaptor A2
A_1
S_i
A_2
S_i
Mix conjugate products, ligate

Fig. 1A

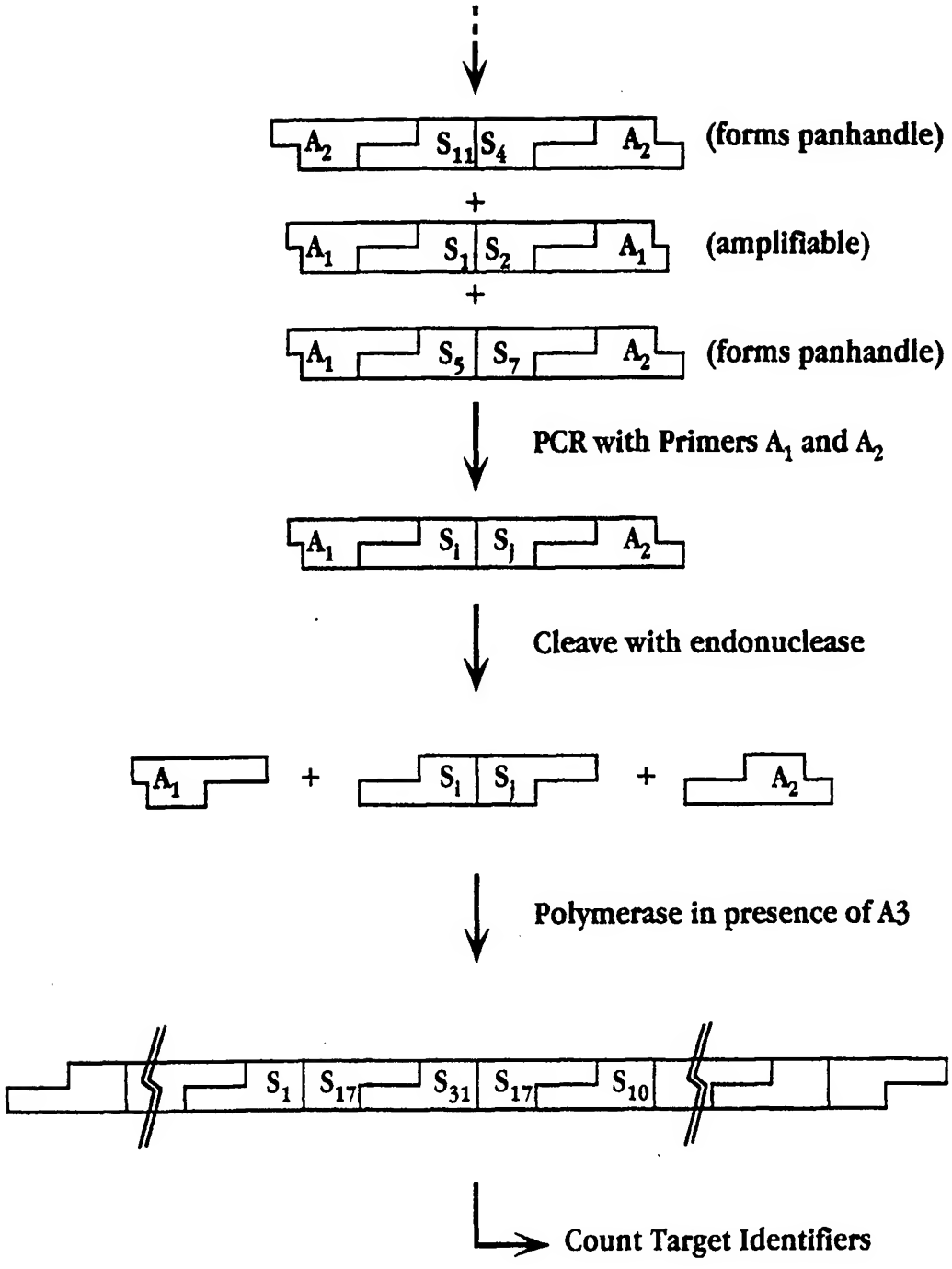

A2 S11 S4 A2
(forms panhandle)
+
A1 S1 S2 A1
(amplifiable)
+
A1 S5 S7 A2
(forms panhandle)
PCR with Primers A1 and A2
A1 Si Sj A2
Cleave with endonuclease
A1 + Si Sj + A2
Polymerase in presence of A3
S1 S17 S31 S17 S10
Count Target Identifiers

Fig. 1B

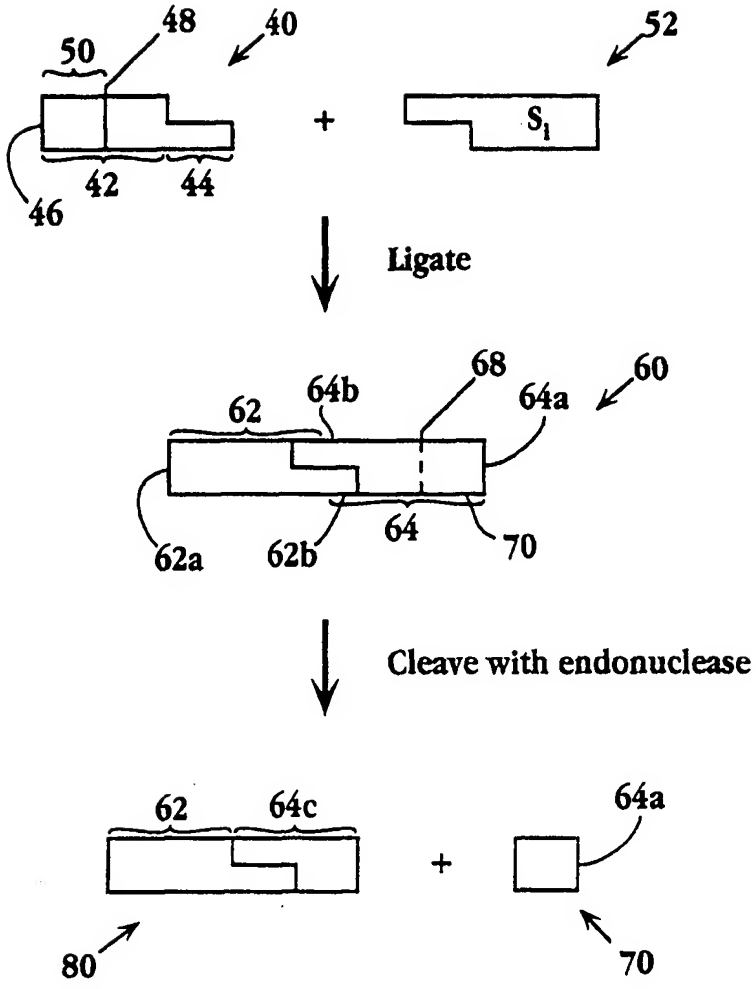

48
40
50
52
+
S_1
46
42
44
Ligate
68
64b
64a
60
62
62a
62b
64
70
Cleave with endonuclease
62
64c
64a
+
80
70

Fig. 2

Adaptor #1:

5′-TTTGGATTTGCTGGTGCAGTACAACTAGGCTTAATA**GGGAC**ATG-OH 3′

3′-CGCCTAAACGACCACGTCATGTTGATCCGAATTAT**CCCT**p-5′

Adaptor #2:

5′-TTTCTGCTCGAATTCAAGCTTCTAACGATGTACG**GGGAC**ATG-OH-3′

3′-CGGACGAGCTTAAGTTCGAAGATTGCTACATGC**CCCT**p-5′

Fig. 3

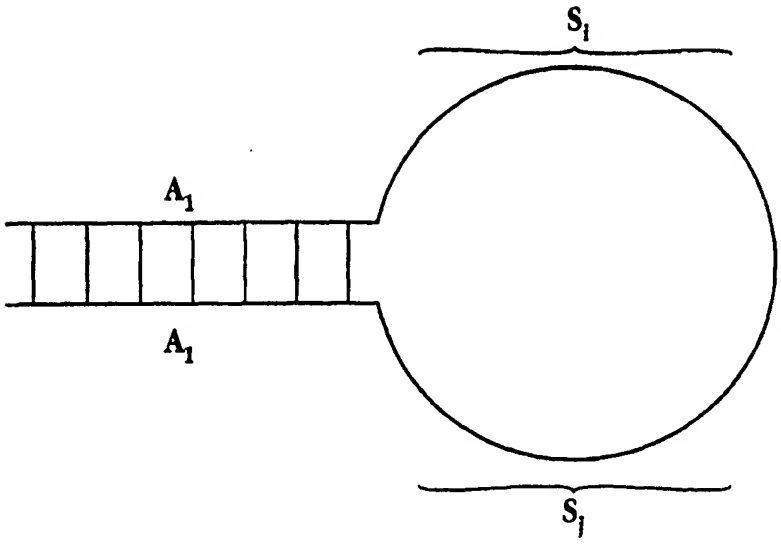


Fig. 4

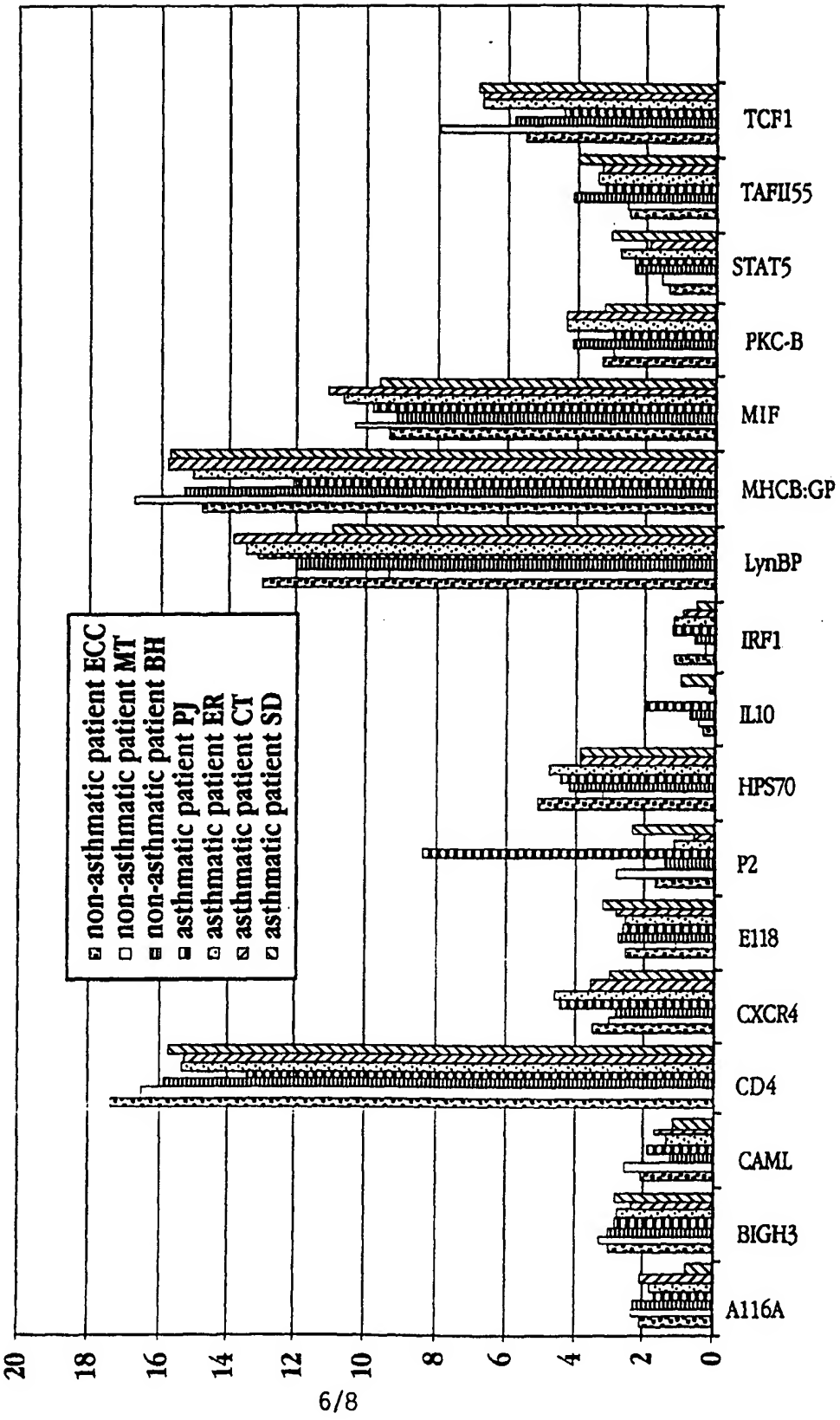

non-asthmatic patient ECC
non-asthmatic patient MT
non-asthmatic patient BH
asthmatic patient PJ
asthmatic patient ER
asthmatic patient CT
asthmatic patient SD
TCF1
TAFII55
STAT5
PKC-B
M1F
MHCB:GP
LynBP
IRF1
IL10
HPS70
P2
E118
CXCR4
CD4
CAML
BIGH3
A116A
20
18
16
14
12
10
8
6
4
2
0

Fig. 5

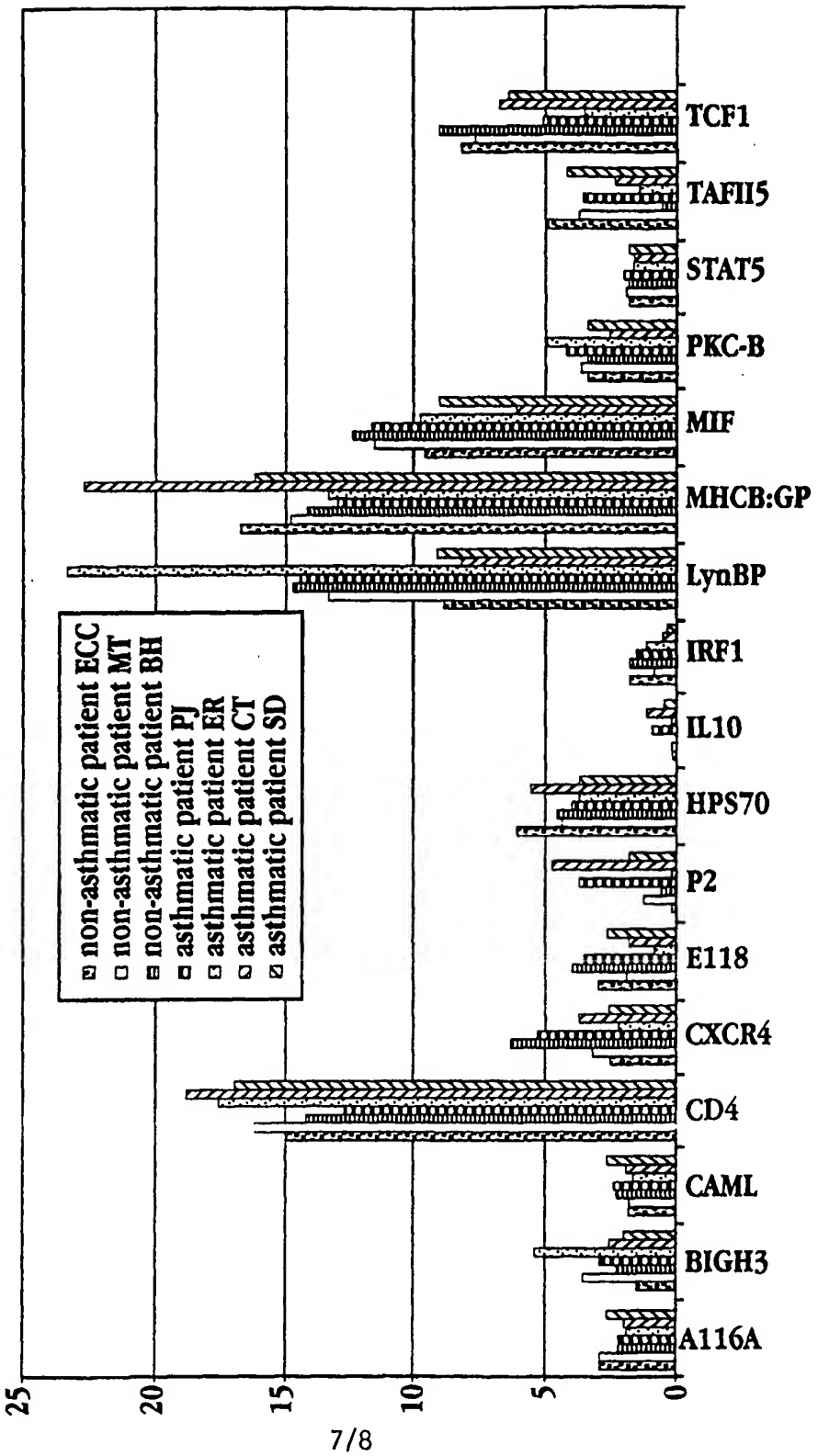

non-asthmatic patient ECC
non-asthmatic patient MT
non-asthmatic patient BH
asthmatic patient PJ
asthmatic patient ER
asthmatic patient CT
asthmatic patient SD
TCF1
TAFII5
STAT5
PKC-B
MIF
MHCB:GP
LynBP
IRF1
IL10
HPS70
P2
E118
CXCR4
CD4
CAML
BIGH3
A116A
25
20
15
10
5
0

Fig. 6

```
human P2      1 MWHLKLCAVLMIFLLLLGQIDGSPIPEVSSAKRRPRRMTPFWRGVSLRPI 50
mouse P2      1 MLQLKLFAVLLTCLLLLGQVNSSPVPEVSSAKRS-RRMTPFWRGVSLRPI 49
                *  *** ***   ******   ** ********  ***************

human P2     51 GASCRDDSECITRLCRKRRCSLSVAQE 77
mouse P2     50 GASCRDDSECITRLCRKRRCSLSVAQE 76
                ****************************
```

Fig. 7

SEQUENCE LISTING

<110> Dolganov, Gregory
Novikov, Alexander

<120> METHOD FOR MEASURING TARGET POLYNUCLEOTIDES, AND NOVEL ASTHMA BIOMOLECULES

<130> G65PCT

<140> unassigned
<141> 1999-11-11

<150> 09/193,320
<151> 1998-11-16

<160> 50

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 1459
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (192)...(422)

<400> 1

```
tacaacgtgc ctgaaaggtg gcccttgagc acattcctcc ggtagacatt aacttattaa     60
attgattctg attacaaata taaactttgc ccccatctca cccagtaaca atgcaagagt    120
tgatgtcagt ctataaaagg aagtaggaac tgtcccctggc tttcaggctc caacatcctc    180
cccctgtcaa g atg tgg cac ctc aaa ctt tgt gca gtc ctc atg atc ttc      230
             Met Trp His Leu Lys Leu Cys Ala Val Leu Met Ile Phe
              1               5                  10

ctg ttg ctg ttg ggc cag ata gat ggc tcc cca ata cca gaa gtg agt       278
Leu Leu Leu Leu Gly Gln Ile Asp Gly Ser Pro Ile Pro Glu Val Ser
     15                  20                  25

tcg gca aag aga agg cca cgg aga atg acc cca ttt tgg aga ggg gtt       326
Ser Ala Lys Arg Arg Pro Arg Arg Met Thr Pro Phe Trp Arg Gly Val
 30                  35                  40                  45

tcc ctc agg cct att gga gcc tcc tgc cgg gat gat tct gag tgt atc       374
Ser Leu Arg Pro Ile Gly Ala Ser Cys Arg Asp Asp Ser Glu Cys Ile
                 50                  55                  60

aca agg cta tgc aga aaa aga cgc tgt tcc tta agt gtg gcc cag gaa       422
Thr Arg Leu Cys Arg Lys Arg Arg Cys Ser Leu Ser Val Ala Gln Glu
             65                  70                  75

tgatgtacat accagggaaa gaaaggacag cagtcacctc cgacaatgct ccgttctatg    482
gaatattgat taactgcatt ttggctggag acacccaagt gaagcaatct tgtattttta    542
atatttaaag gcagatgtac gctttaaatt ggtctccatt tcttcttaga atgttgatat    602
atggataagc ataactaaac ttgtcaattt agagtttatt tttctatgga tactattaaa    662
tgtctcaaat tgaaatttta gcagtctgga attcaagctt ttgagggaaa gaaggattca    722
ttttgtatac taaagaaaaa aacagcattg cccaataatg tgttaacttc tcaatctgga    782
```

```
aagtgtagtg agagctacat aatcaatagc tacgtaatca acttcagcaa gttcctaagc    842
tgtggccctg gatcccttca ctccatactc ttcagggagg tgtcaaaggt ggtcaagctt    902
gggaggctga ggcaggagaa tcacttgaac ctgggaggtg gaggttgcag tgagccaaga    962
tcatgccact gcaccccagc ctgggtgaca tagagagact ctgtctcaaa aaaaaaaaaa   1022
aaaaaaaaag atggtcaagc acttgccatc taatgaagca cccaggtttg tctatggctg   1082
tctgaagata aggcagaaca gagagaaaaa tctagtttaa taaagcaatg aaagcccaag   1142
ccaccaaaag caccaaaaga gatttaagat ttctgttggg acaaggtgaa ttttgacata   1202
tgcacaagga cccatggcca gcaagttcag aatttctgta atcagaaaat gagacctcag   1262
ctggtgggaa cagacttta aagagtagta acggacctta agacacaaaa caggaattgc   1322
atgaaaaggg ctcagagttg aagaaatata aagagtacag tgttcagtaa ctttaataca   1382
gaaaatctat ttgatattta ttaaacttag tttctccaac tatggtttgg cattgtacaa   1442
agcatcttaa tgacaca                                                  1459
```

<210> 2
<211> 77
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Trp His Leu Lys Leu Cys Ala Val Leu Met Ile Phe Leu Leu Leu
 1               5                  10                  15
Leu Gly Gln Ile Asp Gly Ser Pro Ile Pro Glu Val Ser Ser Ala Lys
            20                  25                  30
Arg Arg Pro Arg Arg Met Thr Pro Phe Trp Arg Gly Val Ser Leu Arg
        35                  40                  45
Pro Ile Gly Ala Ser Cys Arg Asp Asp Ser Glu Cys Ile Thr Arg Leu
    50                  55                  60
Cys Arg Lys Arg Arg Cys Ser Leu Ser Val Ala Gln Glu
65                  70                  75
```

<210> 3
<211> 1657
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (192)...(311)

<400> 3

```
tacaacgtgc ctgaaaggtg gcccttgagc acattcctcc ggtagacatt aacttattaa     60
attgattctg attacaaata taaacttttgc ccccatctca cccagtaaca atgcaagagt    120
tgatgtcagt ctataaaagg aagtaggaac tgtccctggc tttcaggctc caacatcctc    180
cccctgtcaa g atg tgg cac ctc aaa ctt tgt gca gtc ctc atg atc ttc      230
             Met Trp His Leu Lys Leu Cys Ala Val Leu Met Ile Phe
              1               5                  10

ctg ttg ctg ttg ggc cag gta agg agg gaa gga tac tta tgt gtg tgt      278
Leu Leu Leu Leu Gly Gln Val Arg Arg Glu Gly Tyr Leu Cys Val Cys
     15                  20                  25

gtg gag tgt gga gat gat agt ggt ggt gga act tgaaagctag attcagtcct    331
Val Glu Cys Gly Asp Asp Ser Gly Gly Gly Thr
 30                  35                  40

gaggaatggt tcctctgttc tgagtctaca gcatctgcgg aatggaatga tcactcttcc    391
aaggtgtgca gcagggtgtc aacactttca tatctgaatg tctttgccct tacagataga    451
tggctcccca ataccagaag tgagttcggc aaagagaagg ccacggagaa tgaccccatt    511
ttggagaggg gtttccctca ggcctattgg agcctcctgc cgggatgatt ctgagtgtat    571
cacaaggcta tgcagaaaaa gacgctgttc cttaagtgtg gcccaggaat gatgtacata    631
ccagggaaag aaaggacagc agtcacctcc gacaatgctc cgttctatgg aatattgatt    691
```

```
aactgcattt tggctggaga cacccaagtg aagcaatctt gtatttttaa tatttaaagg    751
cagatgtacg ctttaaattg gtctccattt cttcttagaa tgttgatata tggataagca    811
taactaaact tgtcaattta gagtttattt ttctatggat actattaaat gtctcaaatt    871
gaaattttag cagtctggaa ttcaagcttt tgagggaaag aaggattcat tttgtatact    931
aaagaaaaaa acagcattgc ccaataatgt gttaacttct caatctggaa agtgtagtga    991
gagctacata atcaatagct acgtaatcaa cttcagcaag ttcctaagct gtggccctgg   1051
atcccttcac tccatactct tcagggaggt gtcaaaggtg gtcaagcttg ggaggctgag   1111
gcaggagaat cacttgaacc tgggaggtgg aggttgcagt gagccaagat catgccactg   1171
caccccagcc tgggtgacat agagagactc tgtctcaaaa aaaaaaaaaa aaaaaaaaga   1231
tggtcaagca cttgccatct aatgaagcac ccaggtttgt ctatggctgt ctgaagataa   1291
ggcagaacag agagaaaaat ctagtttaat aaagcaatga aagcccaagc caccaaaagc   1351
accaaaagag atttaagatt tctgttggga caaggtgaat tttgacatat gcacaaggac   1411
ccatggccag caagttcaga atttctgtaa tcagaaaatg agacctcagc tggtgggaac   1471
agacttttaa agagtagtaa cggaccttaa gacacaaaac aggaattgca tgaaaagggc   1531
tcagagttga agaaatataa agagtacagt gttcagtaac tttaatacag aaaatctatt   1591
tgatatttat taaacttagt ttctccaact atggtttggc attgtacaaa gcatcttaat   1651
gacaca                                                              1657
```

<210> 4
<211> 40
<212> PRT
<213> Homo sapiens

<400> 4

```
Met Trp His Leu Lys Leu Cys Ala Val Leu Met Ile Phe Leu Leu Leu
 1               5                  10                  15
Leu Gly Gln Val Arg Arg Glu Gly Tyr Leu Cys Val Cys Val Glu Cys
            20                  25                  30
Gly Asp Asp Ser Gly Gly Gly Thr
        35                  40
```

<210> 5
<211> 3744
<212> DNA
<213> Homo sapiens

<400> 5

```
ctcccttggc aagaagctcc ctctgtctga ctggcagcag ggtcagtggc catccttttc     60
taccttcagc tgttgtgcat ctcctccacg ttttcctgga gtaaaggaac aagcaggttc    120
ttatccaaat acccaacctg tgaagatgag acggcccttg ggaagaggta ttagccaagt    180
tttcagtatg atgcccggca gaaagaattt gttgtctttg tcttaagctt attatcctcc    240
caggtcatag gaagatctaa agtccaatca gtctccttac tcagaattca cagctttccc    300
ctcaaactta aaatgcccat gtcatttcac gagtacccat aactacccat cgtttagata    360
catgtcctcc cttggtctag aaatctgcag cagttcctac ttaaactgat ggatgctttg    420
ggctttggag gtatctcatc cctatgcctg ttgccctgaa gctgccggaa ggaaaggccc    480
cactcaacta ctgggaagaa agcccataat tcaaggcttt taatacagtc actacacaag    540
ctctccctct gatttaggag atcttgttac ctcctcatct acatgtgctt aaaaacaaaa    600
agaaacctat aattattcaa ttcacctgtg gtccaaataa gtctcacatc cagccattcc    660
cctgagaggg caggaggcca gactttgaac tttccacagc tgggtcttca ttttactagt    720
gagataaaga ggtgcagatt agggtgacag tccatctaag caagtggacg atctggggaa    780
ctgtaggcct tgctgaagca tcctcccaca gaaaggagat aagggcctta tccaattggc    840
tgtacacaat agatcagtgc ttctttccta gacaaggctg aaaggggcca acattatttc    900
tgaagacttc attattggaa ttctatggga gtgatctcac tgagctattt tggaatagaa    960
atgtgtggct agttgcctga cctccctcaa tggtttcacg tggctttcaa agggaaggaa   1020
gggcagtgct gacttttggt aaaatgggcg aaagggtcca tgccagcaac acaatcactc   1080
aaagtccaga tgagggatca gtaaatacaa cgtgcctgaa aggtggccct tgagcacatt   1140
cctccggtag acattaactt attaaattga ttctgattac aaatataaac tttgccccca   1200
tctcacccag taacaatgca agagttgatg tcagtctata aaaggaagta ggaactgtcc   1260
ctggctttca ggctccaaca tcctcccccc gtcaagatgt ggcacctcaa actttgtgca   1320
gtcctcatga tcttcctgtt gctgttgggc caggtaagga gggaaggata cttatgtgtg   1380
tgtgtggagt gtggagatga tagtggtggt ggaacttgaa agctagattc agtcctgagg   1440
```

```
aatggttcct ctgttctgag tctacagcat ctgcggaatg gaatgatcac tcttccaagg    1500
tgtgcagcag ggtgtcaaca ctttcatatc tgaatgtctt tgcccttaca gatagatggc    1560
tccccaatac cagaagtgag ttcggcaaag agaaggccac ggagaatgac cccattttgg    1620
agaggggttt ccctcaggcc tattggagcc tcctgccggg atgattctga gtgtatcaca    1680
aggctatgca ggtactccct gaacctggga gcagggttgg gccagagagc cctgggaagc    1740
tgggcagagg agtgacaagg ggacacatga acctaagaat aaagctggga tggaggagtt    1800
ctagactgag actgggagct ccatggaaaa tcccttagga gttaggagcc aagaacagct    1860
ctatgtgggt accatgaaag agtggtgcct ttccattctt cagtctttcc tgagactgga    1920
cagatcaagc aaagaggaag gaaatgcagc ataggtggcc ctggtcattc ctagacccag    1980
tcttaaaaaa ctacccttte ttttccccta gaaaaagacg ctgttcctta agtgtggccc    2040
aggaatgatg tacataccag ggaaagaaag gacagcagtc acctccgaca atgctccgtt    2100
ctatggaata ttgattaact gcattttggc tggagacacc caagtgaagc aatcttgtat    2160
ttttaatatt taaaggcaga tgtacgcttt aaattggtct ccatttcttc ttagaatgtt    2220
gatatatgga taagcataac taaacttgtc aatttagagt ttatttttct atggatacta    2280
ttaaatgtct caaattgaaa ttttagcagt ctggaattca agcttttgag ggaaagaagg    2340
attcattttg tatactaaag aaaaaaacag cattgcccaa taatgtgtta acttctcaat    2400
ctggaaagtg tagtgagagc tacataatca atagctacgt aatcaacttc agcaagttcc    2460
taagctgtgg ccctggatcc cttcactcca tactcttcag ggaggtgtca aaggtggtca    2520
agcttgggag gctgaggcag gagaatcact tgaacctggg aggtggaggt tgcagtgagc    2580
caagatcatg ccactgcacc ccagcctggg tgacatagag agactctgtc tcaaaaaaaa    2640
aaaaaaaaaa aaaagatggt caagcacttg ccatctaatg aagcacccag gtttgtctat    2700
ggctgtctga agataaggca gaacagagag aaaaatctag tttaataaag caatgaaagc    2760
ccaagccacc aaaagcacca aaagagattt aagatttctg ttgggacaag gtgaattttg    2820
acatatgcac aaggacccat ggccagcaag ttcagaattt ctgtaatcag aaaatgagac    2880
ctcagctggt gggaacagac ttttaaagag tagtaacgga ccttaagaca caaaacagga    2940
attgcatgaa aagggctcag agttgaagaa atataaagag tacagtgttc agtaacttta    3000
atacagaaaa tctatttgat atttattaaa cttagtttct ccaactatgg tttggcattg    3060
tacaaagcat cttaatgaca cagtagagtt aaaaagatct ttacaaattg aaatgtgata    3120
cccttgtctc caaatatttt aattgccata aatttgttta aaaatacaca ttaaacgcat    3180
gtttgacact ctaaactttc tatactctca ataccacaaa atacatataa tatactatac    3240
agatgtggaa aaatctagtt agtatataat actttgctca ccattaacag ctttatcgtg    3300
tgaaatgcac atactgactt agaaatccta gcttttgagc cccaaagtac ctctgaaatt    3360
gtaatgtatt gcaacaaata aaaatcacag tatataattg aaattctggt tgaaaatgtc    3420
agatgcctaa atattttgga taaaagaaat gtaaaacaaa gatttggttc atgtacaaaa    3480
caaaggcatc cttatcagtc actaacagta ccttttctgca aaatcaacag gcttttaatt    3540
tatcagtgac attatctccc ctccaaaacc ctccccaaat atcaaacaat tatataccaa    3600
ataattttaa tttcaaataa aaggattttc aataaatata taagcatttc catcctccat    3660
atacaaaatt tcttaaaacc attcaaaacg gaggctagat agaaattttc ataactgtgc    3720
ttaccaaccc cacaactggc cctt                                           3744
```

<210> 6
<211> 651
<212> DNA
<213> Mus

<220>
<221> CDS
<222> (69)...(299)

<400> 6

```
gtcggtctat aaaaggaaac gggaactgcc ccctggcttt cagactccaa catcattctt     60
ttgccaag atg cta cag cta aaa ctc ttt gca gtg ctc ctg act tgc ctg     110
         Met Leu Gln Leu Lys Leu Phe Ala Val Leu Leu Thr Cys Leu
          1               5                  10

ctg ctg ctg ggt cag gtc aat agt tcc cca gta cca gaa gtg agt tca      158
Leu Leu Leu Gly Gln Val Asn Ser Ser Pro Val Pro Glu Val Ser Ser
 15                  20                  25                  30
```

```
gca aag aga tcc cgg aga atg acc cca ttt tgg aga ggg gtt tcc ctc      206
Ala Lys Arg Ser Arg Arg Met Thr Pro Phe Trp Arg Gly Val Ser Leu
                  35                  40                  45

agg ccc att ggt gcc tcg tgc cgg gat gac tct gag tgt atc aca aga      254
Arg Pro Ile Gly Ala Ser Cys Arg Asp Asp Ser Glu Cys Ile Thr Arg
              50                  55                  60

cta tgc aga aaa aga cgc tgt tcc cta agt gtg gcc cag gag tga          299
Leu Cys Arg Lys Arg Arg Cys Ser Leu Ser Val Ala Gln Glu  *
          65                  70                  75

tgtccatact ggggaagagt cacctccaac agtgatctgt tctggggaat attgttaacg    359
acattttggt tggagataga taagtgaagt agtctttgtt tttaatattt aaagacggat    419
actttgcttta aactggtctc catttcttat tagaatgttg ttgatttgta catagacata    479
actttgtcag agtttatttt tctataaata ctattaaatg tctcaaatct aagtttctggc    539
agtctggaat tgacacttca ctgagtcact tcatccatta tagggagtag tagtggctcc    599
ctaaccagtg gtgctgctcc acctgggagt gtggtgagca cccactcctc aa            651
```

<210> 7
<211> 76
<212> PRT
<213> Mus

<400> 7

```
Met Leu Gln Leu Lys Leu Phe Ala Val Leu Leu Thr Cys Leu Leu Leu
 1               5                  10                  15
Leu Gly Gln Val Asn Ser Ser Pro Val Pro Glu Val Ser Ser Ala Lys
            20                  25                  30
Arg Ser Arg Arg Met Thr Pro Phe Trp Arg Gly Val Ser Leu Arg Pro
        35                  40                  45
Ile Gly Ala Ser Cys Arg Asp Asp Ser Glu Cys Ile Thr Arg Leu Cys
    50                  55                  60
Arg Lys Arg Arg Cys Ser Leu Ser Val Ala Gln Glu
65                  70                  75
```

<210> 8
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Adaptor 1, forward strand

<400> 8

```
tttggatttg ctggtgcagt acaactaggc ttaataggga catg                    44
```

<210> 9
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Adaptor 1, reverse strand

<400> 9

```
tccctattaa gcctagttgt actgcaccag caaatccgc                          39
```

<210> 10
<211> 42
<212> DNA

<213> Artificial Sequence

<220>
<223> Adaptor 2, forward strand

<400> 10
tttctgctcg aattcaagct tctaacgatg tacggggaca tg 42

<210> 11
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Adaptor 2, reverse strand

<400> 11
tccccgtaca tcgttagaag cttgaattcg agcaggc 37

<210> 12
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Example 4, forward primer

<400> 12
aggcttaata gggacat 17

<210> 13
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Example 4, reverse primer

<400> 13
acgatgtacg gggacat 17

<210> 14
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Example 4, primer PDS1-CUA

<400> 14
cuacuacuac uactgagtat ctctaggatc ctctc 35

<210> 15
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Adaptor AD3, forward strand

<400> 15

tttctactga gtatctctag gatcctctca tg 32

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Adaptor AD3, reverse strand

<400> 16
agaggatcct agagatactc agtag 25

<210> 17
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer A116A-1

<400> 17
gaaatcatgt ctgagggtgt acgtttga 28

<210> 18
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer A116A-2

<400> 18
cttcctggct tggatcttct ccaatctg 28

<210> 19
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer BIGH3-1

<400> 19
gcgagtcctg actgatgagc tgaaacac 28

<210> 20
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> primer BIGH3-2

<400> 20
ggaattctgg tacatagagg tgagggt 27

<210> 21
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> primer CAML-1

<400> 21
gtcatgaact gcttgattat tggggctct 29

<210> 22
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> primer CAML-2

<400> 22
gcttctcctt ctcagttcta caggctt 27

<210> 23
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> primer CD4-1

<400> 23
tgccagtgcc ctcaccggtt t 21

<210> 24
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> primer CD4-2

<400> 24
tgcctggcct cgtgcctcaa a 21

<210> 25
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer CXCR4-1

<400> 25
ccgaggaaat gggctcaggg gactatga 28

<210> 26
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> primer CXCR4-2

<400> 26
tagcattttc ttcacggaaa cagggttcct t 31

<210> 27
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> primer E118-1

<400> 27
gcagtaacat ctgattctaa cagcacatc 29

<210> 28
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer E118-2

<400> 28
cattgaatct ctgaaggctc tatgaaag 28

<210> 29
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer P2-1

<400> 29
acctgggagg tggaggttgc agtg 24

<210> 30
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer P2-2

<400> 30
tatgtcaccc aggctggggt gcag 24

<210> 31
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer HSP70-1

<400> 31
agaagatgca agtggaccag gaggaacc 28

<210> 32
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer HSP70-2

<400> 32
ctggtgtctg ctgctgttgc tcttcaac 28

<210> 33
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer IL10-1

<400> 33
cccaggagat ggaagttgca gtgagctg 28

<210> 34
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer IL10-2

<400> 34
ttgctctgtc acccaggctg gagtacag 28

<210> 35
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer IRF1-1

<400> 35
gggctgagtc tacagcgtgt cttc 24

<210> 36
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer IRF1-2

<400> 36
aggctgtcca gccaggtggc atcc 24

<210> 37
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> primer LYNB-1

<400> 37
agggaggagc ccatttacat catcacc 27

<210> 38
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer LYNB-2

<400> 38
cgctcttcag gaaatccagc aaactgcc 28

<210> 39
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer MHCB:GP-1

<400> 39
tccctctgtg cttctggagg caaggatg 28

<210> 40
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer MHCB:GP-2

<400> 40
aggtgtttgc cttcgttgag atcccata 28

<210> 41
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer MIF-1

<400> 41
gcccggacag ggtctacatc aact 24

<210> 42
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer MIF-2

<400> 42
tggagttgtt ccagcccaca ctgg 24

<210> 43
<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> primer PKCI-1

<400> 43
tgaaggttca gatggtggac agtc 24

<210> 44
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer PKCI-2

<400> 44
atgcatttgc cgacctccaa gaac 24

<210> 45
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> primer STAT5-1

<400> 45
cagtgaaaga caagctgcaa atcagccatg 30

<210> 46
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer STAT5-2

<400> 46
ttttacccct cactccagat tccacagg 28

<210> 47
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> primer TAFII55-1

<400> 47
ctgacaattg agttacatcc tgatggg 27

<210> 48
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> primer TAFII55-2

<400> 48
caatggaaca cggtccactc tgacgatt 28

<210> 49

<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer TCF1-1

<400> 49
agtactatga gctggcccgc aagg 24

<210> 50
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> primer TCF1-2

<400> 50
gctgaccagc ctgggtatag ctgc 24

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/63, C07K 14/52, 16/24, C12Q 1/68 — A3

(11) International Publication Number: **WO 00/29621**

(43) International Publication Date: 25 May 2000 (25.05.00)

(21) International Application Number: PCT/US99/26931

(22) International Filing Date: 12 November 1999 (12.11.99)

(30) Priority Data:
09/193,320 16 November 1998 (16.11.98) US

(71) Applicant: GENELABS TECHNOLOGIES, INC. [US/US]; 505 Penobscot Drive, Redwood City, CA 94063–4738 (US).

(72) Inventors: DOLGANOV, Gregory; 14 Chicory Lane, San Carlos, CA 94070 (US). NOVIKOV, Alexander; 5866 Carmel Way, Union City, CA 94587 (US).

(74) Agents: SUH, Young, J.; Legal Department, Genelabs Technologies, Inc., 505 Penobscot Drive, Redwood City, CA 94063–4738 (US) et al.

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

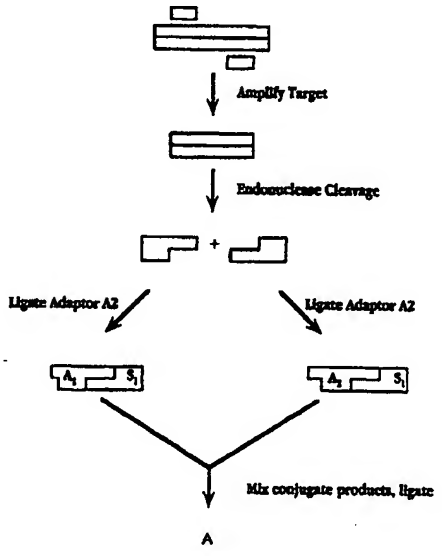
Published
With international search report.

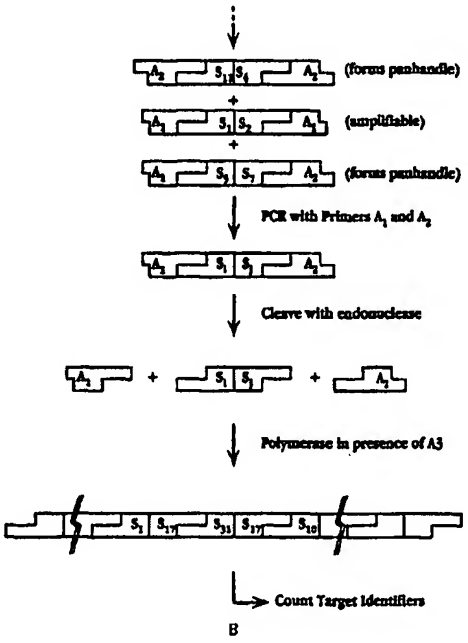
(88) Date of publication of the international search report:
9 November 2000 (09.11.00)

(54) Title: METHOD FOR MEASURING TARGET POLYNUCLEOTIDES AND NOVEL ASTHMA BIOMOLECULES

(57) **Abstract**

The present invention provides a method for simultaneously determining the levels of selected target polynucleotide sequences in a sample. In the method, selected target sequences are amplified using sequence–selective primer pairs to form double–stranded copies. The copies are cleaved with one or more endonucleases, and first and second aliquots of the cleaved fragment mixture are ligated to first and second adaptors, respectively, to form mixtures of first and second adaptor–fragment conjugates. The aliquots are combined and ligated to form conjugate dimers containing a conjugate from each of the first and second conjugate mixtures. After optional amplification, the conjugates are treated to release the adaptor segments, yielding target–identifier dimers. The dimers are polymerized to form target–identifier dimer multimers, and the relative abundances of the target–identifiers in one or more dimer multimers to provide an estimate of the levels of the selected target sequences in the sample. The present invention is particularly useful for determining expression levels of mRNA gene sequences in a sample, and for evaluating changes in mRNA expression levels among different samples or in response to changes in sample conditions. Also disclosed are certain polynucleotides and polypeptides that are useful for a variety of applications, particularly relating to asthma.





FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/US 99/26931

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/63 C07K14/52 C07K16/24 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
WPI Data, PAJ, MEDLINE, CHEM ABS Data, BIOSIS, EMBASE, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 37191 A (SCHULZ KNAPPE PETER ;NEHLS MICHAEL (DE); FORSSMANN WOLF GEORG (DE)) 27 August 1998 (1998-08-27) the whole document	28-50
X	DATABASE EMBL [Online] AC/ID: AC004500, 6 April 1998 (1998-04-06) KIMMERLY ET AL.: "HOMO SAPIENS CHROMOSOME 5, P1 CLONE 107689" XP002141569 abstract	31-39
Y	WO 97 10363 A (UNIV JOHNS HOPKINS MED) 20 March 1997 (1997-03-20) the whole document	1-27

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 August 2000	14. 08. 2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hagenmaier, S

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Intern ial Application No
PCT/US 99/26931

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	VELCULESCU V E ET AL: "SERIAL ANALYSIS OF GENE EXPRESSION" SCIENCE,US,AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, vol. 270, 20 October 1995 (1995-10-20), pages 484-487, XP002053721 ISSN: 0036-8075 the whole document	1-27
Y	EP 0 743 367 A (MAX PLANCK GESELLSCHAFT) 20 November 1996 (1996-11-20) the whole document	1-27
Y	WO 96 34985 A (SUGEN INC) 7 November 1996 (1996-11-07) the whole document	1-27
Y	STONE B AND WHARTON W: "Targeted RNA fingerprinting: The cloning of differentially-expressed cDNA fragments enriched for members of the zinc finger gene family" NUCLEIC ACIDS RESEARCH,GB,OXFORD UNIVERSITY PRESS, SURREY, vol. 22, no. 13, 1 January 1994 (1994-01-01), pages 2612-2618, XP002090069 ISSN: 0305-1048 the whole document	1-27
Y	WO 94 23023 A (UNIV ROCHESTER) 13 October 1994 (1994-10-13) the whole document	1-27
Y	WO 97 22720 A (BEATTIE KENNETH LOREN) 26 June 1997 (1997-06-26) the whole document	1-6, 13-17
A	WO 93 18176 A (DANA FARBER CANCER INST INC) 16 September 1993 (1993-09-16) the whole document	
A	WO 98 10095 A (BRAX GENOMICS LTD ;THOMPSON ANDREW HUGIN (GB); SCHMIDT GUNTER (GB)) 12 March 1998 (1998-03-12) the whole document	
A	COUSINS D J ET AL : "Regulation of cytokine genes implicated in asthma and atopy " GENETICS OF ASTHMA AND ATOPY, vol. 33, November 1996 (1996-11), pages 138-152, XP002101568 the whole document	

-/--

INTERNATIONAL SEARCH REPORT

Intern ial Application No
PCT/US 99/26931

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 27284 A (GENELABS TECH INC ;DOLGANOV GREGORY (US)) 31 July 1997 (1997-07-31) the whole document	

INTERNATIONAL SEARCH REPORT

Int. ational application No.
PCT/US 99/26931

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☒ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☒ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-27

Method for simultaneously determining the levels of a plurality of selected target gene sequences in a sample by amplifying the selected target sequences using sequence-selective primer-pairs to form double-stranded copies which are cleaved with one or more endonucleases, and first and second aliquots of the cleaved fragment mixture are ligated to first and second adaptors, respectively to form mixtures of first and second adaptor-fragment conjugates which are combined and ligated to form conjugate dimers containing a conjugate from each of the first and second conjugate mixtures. After optional amplification, the conjugates are treated to release the adaptor segments, yielding target-identifier dimers which are polymerized to form target-identifier dimer multimers to provide an estimate of the levels of the selected target sequences in sample.

2. Claims: 28-50

Isolated and purified polypeptides with at least 80% identity to Seq.ID 2 or 4 and polynucleotides with at least 80% identity to Seq.ID 1 or 3, as well as methods and compositions relating thereto and a polynucleotide selected from the group of
a) nucleotides 1047-1353 of Seq.ID 5
b) nucleotides 1354-1551 of Seq.ID 5
c) nucleotides 1552-1691 of Seq.ID 5
d) nucleotides 1692-2011 of Seq.ID 5
e) nucleotides 2012-2633 of Seq.ID 5
and the complement of any of a)-e).

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter ɔnal Application No
PCT/US 99/26931

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9837191 A	27-08-1998	DE 19730786 A	21-01-1999
WO 9710363 A	20-03-1997	US 5695937 A	09-12-1997
		US 5866330 A	02-02-1999
		AU 707846 B	22-07-1999
		AU 6561496 A	20-03-1997
		AU 7018896 A	01-04-1997
		CA 2185379 A	13-03-1997
		EP 0761822 A	12-03-1997
		GB 2305241 A,B	02-04-1997
		IE 80465 B	12-08-1998
		JP 10511002 T	27-10-1998
EP 0743367 A	20-11-1996	DE 19518505 A	21-11-1996
		JP 8308598 A	26-11-1996
		US 5876932 A	02-03-1999
WO 9634985 A	07-11-1996	US 5830648 A	03-11-1998
		AU 5636996 A	21-11-1996
		CA 2220300 A	07-11-1996
		EP 0824597 A	25-02-1998
		JP 11505420 T	21-05-1999
WO 9423023 A	13-10-1994	US 5643765 A	01-07-1997
		US 5639606 A	17-06-1997
		AU 6493194 A	24-10-1994
		JP 7132100 A	23-05-1995
		US 5876978 A	02-03-1999
WO 9722720 A	26-06-1997	AU 1687597 A	14-07-1997
WO 9318176 A	16-09-1993	US 5262311 A	16-11-1993
		CA 2102784 A	12-09-1993
		EP 0592626 A	20-04-1994
		JP 2843675 B	06-01-1999
		JP 7500735 T	26-01-1995
		US 5599672 A	04-02-1997
		US 5665547 A	09-09-1997
		US 5965409 A	12-10-1999
WO 9810095 A	12-03-1998	AU 4027497 A	26-03-1998
		CN 1234076 A	03-11-1999
		EP 0927267 A	07-07-1999
WO 9727284 A	31-07-1997	US 5821091 A	13-10-1998
		US 5965427 A	12-10-1999
		AU 1842597 A	20-08-1997